

Zur Qualitätsbeurteilung von 3D-Videoobjekten

Dissertation

zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

vorgelegt der Fakultät für Elektrotechnik und Informationstechnik
der Technischen Universität Ilmenau

von Dipl.-Ing. Marco Rittermann
geboren am 20. April 1972 in Gotha

Gutachter: 1. Univ.-Prof. Dr.-Ing. Karlheinz Brandenburg
Technische Universität Ilmenau
2. Univ.-Prof. Dr.-Ing. Thomas Sikora
Technische Universität Berlin
3. Dr.-Ing. Aljoscha Smolić
Fraunhofer Institut für Nachrichtentechnik HHI, Berlin

Tag der Einreichung: 4. Oktober 2004
Tag der wissenschaftlichen Aussprache: 23. März 2007

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit der Qualitätsbewertung von natürlichen dreidimensionalen Videoobjekten. Dieser neuartige Medienobjekttyp erlaubt die Betrachtung von natürlichen Objekten (z. B. einer Person) aus verschiedenen Perspektiven, die im Idealfall frei wählbar sind. Zunächst werden in dieser Arbeit die Verfahren zur objektiven und subjektiven Qualitätsbewertung von konventionellem Bewegtbild untersucht und systematisiert. Ebenso werden die vielfältigen Verfahren zur Generierung von 3D-Videoobjekten analysiert. Diese bilden ein großes Spektrum von bild- bis hin zu modellbasierten Verfahren. Auf der Grundlage dieser Analyse erfolgt eine Begriffsbestimmung und die Beschreibung eines Modells der 3D-Videoobjektgenerierung, welches für sämtliche Generierungsverfahren gültig ist. Um die Qualität von 3D-Videoobjekten zu untersuchen, wird zunächst die Referenzfrage gelöst. Als Referenzen für zu bewertende 3D-Videoobjekte dienen ausreichend gute 3D-Videoobjekte, welche denselben Inhalt darstellen. Im Weiteren werden die speziellen Qualitätsaspekte von 3D-Videoobjekten aufgezeigt, u. a. Größenfehler, Okklusionen, Epipolarverzerrungen, Kompressionsartefakte und Blickwinkelfehler. Verschiedene Verfahren wie beispielsweise das hierarchische Block-Matching und die eindimensionale DFT dienen dazu, diese Merkmale mittels mathematischer Modelle zu beschreiben. Die wichtigste Einschränkung bei der Auswahl und der Anwendung dieser Methoden ist es, dass a priori kein Bildpunktbezug vorausgesetzt werden kann. Die Merkmalsmodelle umfassen auch die dynamischen Eigenschaften, welche sowohl zeit- als auch blickpunktsänderungsabhängig sein können. Im Anschluss wird die Entwicklung von 3D-Videoobjektqualitätsmerkmalen auf Basis der mathematischen Merkmalsmodelle beschrieben. Diese Qualitätsparameter stellen einen Bezug zur visuellen Wahrnehmung der Qualitätsaspekte und Störungen dar. Es wird eine Methodik zur Bildung eines 3D-Videoobjektqualitätsmaßes $3DVQM$ auf Basis dieser Qualitätsparameter vorgestellt. Die Bestimmung der Gewichtungskoeffizienten erfolgt sukzessive mit Hilfe erweiterter subjektiver Bewertungsverfahren. Zum Schluss wird die vorgeschlagene Methodik exemplarisch für eine Testreihe durchgeführt. In der Auswertung werden die Möglichkeiten der objektiven Bewertung zur Prädiktion der subjektiven Qualitätsbewertung dargelegt.

Abstract

This thesis is focussed on the quality assessment of natural three-dimensional video objects. This novel kind of media objects allows watching a natural object (e. g. an actor) from different viewpoints, ideally from every arbitrary viewpoint. The quality assessment of such objects raises several questions to be solved. Many customary methods are not suitable for this new purpose. Firstly, several methods for the subjective and objective quality assessment of conventional video are investigated and systematized. Moreover, the manifold methods of generating 3D video objects are analysed. They cover a wide range from image-based to model-based methods. This analysis makes it possible to define the term *3D video object* and to develop a model of its process. This model is applicable to all kinds of 3D video objects. In order to assess the quality of a 3D video object the definition of a reference is necessary. A 3D video object showing a sufficient quality has to serve as reference for the 3D video object in evaluation. Furthermore, specific quality aspects of 3D video objects are pointed out, e. g., occlusions, distortions along epipolar lines, artifacts caused by compression, and a differing angle of view. Several methods like hierarchical block matching and a one-dimensional DFT are used to create mathematical models describing these quality features. There are significant restrictions for these methods. Naturally, a 3D video object can not be compared with its reference pixel by pixel. The models describing the quality features cover dynamic properties, too. They are depending on time as well as on the changing of the viewpoint. A set of quality parameters is developed using these mathematical feature models. These quality parameters consider the visual perception of the quality aspects and impairments. A methodology creating a 3D video quality metric *3DVQM* based on these quality parameters is introduced. The parameter weights are determined by a successive approximation using extended subjective quality judgements. Finally, the proposed methodology is carried out exemplary by assessing a set of test objects. With this, the prediction of the subjective assessment based on the proposed objective method is evaluated.

Danksagung

Die vorliegende Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Medientechnik der Technischen Universität Ilmenau. Mein Dank gilt den Mitarbeitern und Studierenden am Institut für Medientechnik, welche mir die Erstellung der vorliegenden Dissertation ermöglichten. Mein besonderer Dank gilt meinem Doktorvater Herrn Univ.-Prof. Dr.-Ing. Karlheinz Brandenburg und den beiden Korreferenten Herrn Univ.-Prof. Dr.-Ing. Thomas Sikora und Herrn Dr.-Ing. Aljoscha Smolić. Weiterhin gilt mein Dank Herrn Dipl.-Ing. Christian Weigel und Herrn Dipl.-Ing. Michael Schuldt für die Hinweise zur vorliegenden Arbeit.

Mein persönlicher Dank gilt meinen Eltern und meiner Frau Rita.

Marco Rittermann

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Zielsetzung	1
1.3	Gliederung	2
2	Bewertung von 2D-Bewegt看	3
2.1	Die Kette von der Akquisition bis zur Visualisierung von 2D-Bewegt看 . .	3
2.2	Qualitätsbestimmende Parameter	4
2.2.1	Parameter der Aufnahme	4
2.2.2	Temporale und lokale Auflösung	5
2.2.3	Luminanz	8
2.2.4	Farbrepräsentation	9
2.2.5	Bandbreite und Übertragungsfehler	10
2.3	Subjektive Bewertung von 2D-Bewegt看	11
2.3.1	Anzahl und Auswahl der Betrachter	11
2.3.2	Testbedingungen	12
2.3.3	Auswahl des Testmaterials	13
2.3.4	Single-Stimulus-Methoden	13
2.3.5	Double-Stimulus-Methoden	14
2.3.6	Objektbasierte Bewertung	16
2.3.7	Analyse von subjektiven Bewertungsergebnissen	16
2.4	Objektive Maße zur Bewertung von 2D-Bewegt看	18
2.4.1	Signaltechnische Maße	18
2.4.2	Merkmalsbasierte Maße	19
2.4.3	HVS-basierte Maße	24
2.4.4	Bewertung der Segmentierungsqualität	27
2.4.5	Systematisierung der Bewertungsmaße	29
3	3D-Videoobjekte und deren Generierung	32
3.1	Begriffseingrenzung	32
3.2	Bildbasierte Verfahren	34
3.2.1	Die Geometrie einer multiskopischen Ansicht	34
3.2.2	Gewinnung von Objektinformationen	37
3.2.3	View-Synthesis	38
3.3	Modellbasierte Verfahren	42
3.3.1	Aufnahme	43
3.3.2	Extraktion eines 3D-Modells	43
3.3.3	Texturierung	44
3.3.4	Voxel-Colouring	44
3.3.5	Relief-Texture-Mapping	45

3.4	Überblick über die Verfahren	45
3.5	Modell der 3D-Videoobjektgenerierung	47
4	Modell der Qualität von 3D-Videoobjekten	48
4.1	Zielstellung für die Modellbildung	48
4.2	Bisherige Arbeiten zur Qualitätsbewertung von dreidimensionalem Video	49
4.3	Referenzen für die 3DVO-Qualitätsbewertung	50
4.3.1	Vergleich von 3D-Videoobjekten	50
4.3.2	Verwendung zusätzlicher Ansichten	52
4.3.3	Grenzen der Methode	52
4.4	Qualitätsmerkmale von 3D-Videoobjekten	53
4.4.1	Perspektivenfehler	53
4.4.2	Silhouettenfehler	53
4.4.3	Synthespezifische Fehler	54
4.4.4	Videospezifische Fehler	55
4.4.5	Verdeckungsfehler (Okklusionen)	55
4.4.6	Dynamik der Qualitätsmerkmale	55
4.5	Ansatz zur Entwicklung von Merkmalsmodellen	56
4.5.1	Methodik zur objektiven Bewertung der Qualitätsmerkmale	56
4.5.2	Bildung von Merkmalsmaßen	58
4.5.3	Untersuchung der Dynamik von Merkmalsmaßen	59
4.5.4	Auswahl geeigneter Methoden	60
4.6	Entwicklung von Merkmalsmodellen	62
4.6.1	Translations- und Skalierungsfehler	62
4.6.2	Perspektivenfehler des Blickwinkels	67
4.6.3	Störungen der Silhouette	72
4.6.4	Fehler bei der Epipolargeometriebestimmung	77
4.6.5	Texturierungsfehler	77
4.6.6	Verdeckungsfehler	80
4.6.7	Weitere Merkmalsmodelle	81
4.7	Dynamik der Merkmalsmaße	81
4.7.1	Zeitabhängigkeit der Merkmalsmaße	81
4.7.2	Blickpunktsänderungsabhängigkeit	83
5	3D-Videoobjektqualitätsmaß: 3DVQM	86
5.1	Methodik für die Bestimmung der Qualitätsparameter	86
5.1.1	Anwendung der Merkmalsmodelle	86
5.1.2	Abfolge zur Bestimmung der Merkmalsmaße und Qualitätsparameter	86
5.1.3	Regression von Qualitätsparametern	87
5.1.4	Exemplarische Bildung einiger Qualitätsparameter	89
5.2	Bildung eines 3D-Videoobjektqualitätsmaßes	94
5.2.1	Zusammenführung von Qualitätsparametern	94
5.2.2	Methodik zur Bestimmung der Gewichtungskoeffizienten	98
5.3	Exemplarische Bestimmung der Gewichtungskoeffizienten	104
5.3.1	Auswahl der Testobjekte	104
5.3.2	Auswahl der zu bewertenden Störungen	105
5.3.3	Bestimmung der Initialwerte für die Koeffizienten	105
5.3.4	Bewertungsreihen	105

5.3.5	Objektive Bewertung	107
5.3.6	Abbildung der objektiven Bewertung auf die subjektive Bewertung	107
5.3.7	Modifikation der Gewichtungskoeffizienten	112
5.3.8	Auswertung der 3DVQM-Approximation	113
5.4	Validierung der Methodik	114
6	Ergebnisse und Ausblick	116
6.1	Zusammenfassung der Ergebnisse	116
6.2	Künftige Arbeiten	119
6.3	Schlussbetrachtung	120
	Abbildungsverzeichnis	121
	Tabellenverzeichnis	123
	Verzeichnis der verwendeten Symbole	124
	Verzeichnis der verwendeten Abkürzungen	125
	Quellenverzeichnis	127

1 Einleitung

1.1 Motivation

In modernen audiovisuellen Anwendungen werden die Inhalte als einzelne Medienobjekte codiert. Der zeitliche und örtliche Zusammenhang der Objekte wird in einer zwei- oder dreidimensionalen Szenenbeschreibung definiert [ISO98]. Die einzelnen Medienobjekte können unterschiedlicher Art (z. B. geformtes Video, modellierte künstliche Welten, Sprache, menschliche Gesichter) und unterschiedlichen Ursprunges (natürlich, synthetisch) sein. Das objekt- und szenenbasierte Konzept weist neue Leistungsmerkmale wie eine erweiterte Skalierbarkeit und eine vielfältige Interaktivität auf [Dru02]. Dem Nutzer können zusätzliche Ansichten angeboten werden oder er kann individuell durch die dreidimensionale Szene navigieren. Während modellierte 3D-Objekte grundsätzlich hierfür geeignet sind, verfügen konventionelle Videoaufnahmen nicht über die notwendige dritte Dimension. Wenn der Nutzer sich um ein beliebig berandetes Videoobjekt (*shaped video object*) bewegt, erhält er eine falsche Perspektive, da er sich vom ursprünglichen Aufnahmepunkt entfernt. Daher wird ein neuer Medienobjektyp benötigt, der dieser Einschränkung nicht unterliegt: *natürliche 3D-Videoobjekte*. Das Hauptmerkmal solcher 3D-Videoobjekte (3DVO) ist, dass sie von verschiedenen Standpunkten aus betrachtet werden können. Im Idealfall können sie von jedem beliebigen Standpunkt aus betrachtet werden. Für die Aufnahme, Verarbeitung, Repräsentation und Wiedergabe von 3DVO sind mehrere Verfahren in Entwicklung, die sich auch im Ansatz grundsätzlich unterscheiden [Rit03]. Man kann die Verfahren grob in bildbasierte und modellbasierte Verfahren unterteilen.

Die verschiedenen 3DVO-Verfahren befinden sich teilweise in einem frühen Entwicklungsstadium und bedürfen noch entscheidender Verbesserungen. Sollen Anwendungen für 3DVO produziert werden, stellt sich daher die Frage, welches 3DVO-Verfahren für den konkreten Anwendungsfall geeignet ist: Wie kann der gewünschte Blickpunktbereich erzielt werden? Werden die an die Anwendung gestellten Qualitätsanforderungen durch die Wahl des 3DVO-Verfahrens und dessen Parameter erfüllt? Dies führt unmittelbar zur prinzipiellen Fragestellung: Wie wird die Qualität eines 3DVO durch den Menschen bewertet? Es ist offensichtlich, dass die vielfach untersuchte Bewertung von konventionellem, zweidimensionalem Video nur zu einem geringen Anteil für 3DVO angewendet werden kann. Auf Grund verschiedener Repräsentationen sowie perspektivischer und methodenspezifischer Fehler ist die Qualitätsbewertung von 3DVO ein neues Forschungsthema, welches bislang kaum untersucht wurde. Hieraus ergab sich die Motivation zu der vorliegenden Arbeit.

1.2 Zielsetzung

Das Ziel dieser Arbeit ist die Entwicklung einer Methodik zur Erfassung von Qualitätsparametern von natürlichen 3D-Videoobjekten. Dafür ist eine Begriffsbestimmung und ein Vergleich von Verfahren zur Generierung solcher Objekte notwendig. Mit der zu entwickelnden Methodik sollen 3D-Videoobjekte, welche mittels unterschiedlicher Verfahren erzeugt wurden, bewertet werden können. Eine zentrale Fragestellung bei der Bewertung

von 3DVO ist der Bezug zu einer Referenz. Für 3DVO kann es keine Referenz in dem Sinne geben, wie sie für konventionelles Video genutzt wird. Daher ist die Verwendung von Referenzen für 3DVO zu untersuchen und zu bestimmen. Im nächsten Schritt sind die Qualitätsmerkmale von 3DVO zu ermitteln und durch mathematische Modelle zu beschreiben. Im letzten Schritt ist eine Methodik zur Bildung eines 3D-Videoobjektqualitätsmaßes auf Basis der modellierten Qualitätsparameter zu entwickeln und exemplarisch nachzuweisen.

1.3 Gliederung

In Kapitel 2 wird ein systematisierter Überblick über die Qualitätsbewertung von konventionellem Video gegeben. Dieser Überblick wird unter der Maßgabe entwickelt, die Prinzipien und die Methoden der Qualitätsbewertung herauszustellen. Diese Vorgehensweise bildet eine Basis für die Entwicklung einer Methodik für die Qualitätsbewertung von 3DVO. Daher wird zunächst ein Blick auf die gesamte Kette von der Aufnahme bis zur Wiedergabe von Video gegeben. Anschließend werden objektive und subjektive Verfahren zur Qualitätsbewertung untersucht. Zum Abschluss werden Verfahren zur Bewertung von Videosegmentierungen vorgestellt.

In Kapitel 3 wird die zweite Säule für diese Arbeit gesetzt. Es werden die Grundlagen und vielfältigen Verfahren zur Generierung von 3D-Videoobjekten erläutert. Zunächst wird eine Begriffseingrenzung vorgenommen, da es noch keine einheitliche Verwendung des Begriffes 3D-Videoobjekt gibt. Abschließend wird ein Modell der 3D-Videoobjektgenerierung hergeleitet.

Ein Modell der Qualität von 3D-Videoobjekten wird in Kapitel 4 vorgestellt. Hierfür wird zunächst auf die Lösung des Referenzproblems eingegangen. Anschließend werden die Qualitätsmerkmale und Störungen von 3D-Videoobjekten analysiert. Auf Basis geeigneter Methoden werden mathematische Modelle für die determinierten Merkmale entwickelt.

Die Bildung und die Verifikation eines 3D-Videoobjektqualitätsmaßes wird in Kapitel 5 dargestellt. Zunächst wird eine geeignete Methodik zur Bestimmung der verschiedenen Qualitätsparameter bis hin zur Bildung eines Qualitätsmaßes vorgestellt. Die Zusammenführung der entwickelten Qualitätsparameter auf der Grundlage erweiterter subjektiver Bewertungsreihen wird am Beispiel einer Testreihe dargestellt. Abschließend wird der Zusammenhang zwischen der objektiven und der subjektiven Bewertung untersucht und das 3D-Videoobjektqualitätsmaß diskutiert.

In Kapitel 6 werden die Ergebnisse dieser Arbeit zusammengefasst und ein Ausblick für künftige Arbeiten gegeben.

2 Bewertung von 2D-Bewegt看

2.1 Die Kette von der Akquisition bis zur Visualisierung von 2D-Bewegt看

Für die Beantwortung der Frage nach der Qualitätsbewertung von Video wird zunächst der zu Grunde liegende Prozess betrachtet. Gegenstand der Untersuchungen sind Anwendungen, deren Ziel die Wiedergabe von Bewegtbild auf einem Anzeigegerät zur Betrachtung durch einen oder mehrere Menschen ist. Die höchste Instanz für die Bewertung von Bewegtbild ist daher die subjektive Einschätzung durch die nutzenden Menschen. Eine objektive Bewertung muss daher das Ziel haben, der subjektiven Bewertung möglichst nahe zu kommen.

Das Bewegtbild durchläuft von der Aufnahme bis zur Wiedergabe typische Stationen, welche in Abbildung 2.1 dargestellt sind. Ausgangspunkt ist die dreidimensionale Welt, welche visuell aufgenommen werden soll. Am Ende der Kette steht der Visus und die Bewertung durch den Menschen.

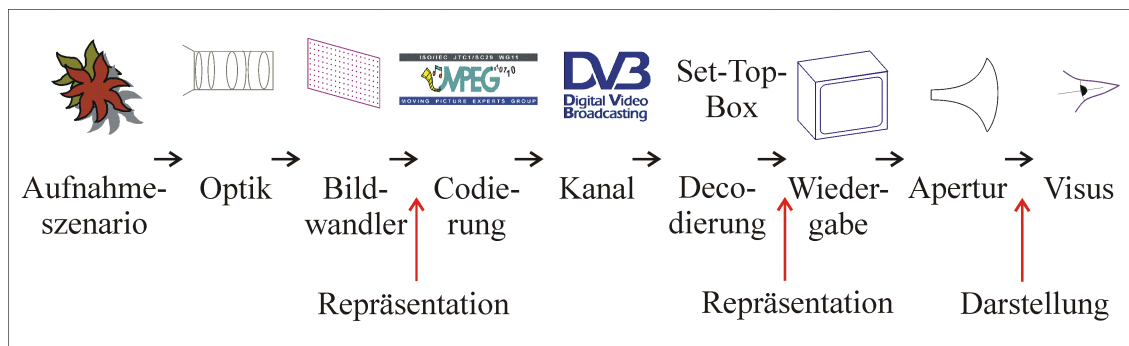


Abbildung 2.1: Kette von der Akquisition bis zur Visualisierung von 2D-Bewegt看

Wenn die Qualität von Video objektiv bewertet wird, so ist in den wenigsten Fällen die gesamte Kette zu bewerten. Soll beispielsweise die Codierung eingeschätzt werden, so wird das decodierte Videosignal mit dem Quellsignal verglichen. Dies ist möglich, da die Repräsentation den gleichen Aufbau und denselben Bezug hat. Im Falle von digitalem Standardfernsehen wird weltweit eine Repräsentation nach ITU-R BT.601 [ITU94a] verwendet, sodass ein einfacher bildpunktweiser Vergleich möglich ist.

Die einzelnen Abschnitte der Kette nehmen unterschiedlichen Einfluss auf die Qualität. Im folgenden Abschnitt werden die qualitätsbestimmenden Parameter der einzelnen Abschnitte dargestellt.

2.2 Qualitätsbestimmende Parameter

2.2.1 Parameter der Aufnahme

Optische Verzerrungen der Linsenoptik

Jede Kameraoptik weist im Gegensatz zur idealen Lochkamera Verzerrungen auf. Zunächst erzeugt jede einzelne Linse Fehler. Typische Fehler einer sphärischen Linse sind [Wen88]:

- *sphärische Aberration*: Zur Linsenmitte hin nimmt die Brennweite zu.
- *chromatische Aberration*: Die Frequenzabhängigkeit der Brechzahl bewirkt unterschiedliche Brennweiten für die einzelnen Farben.
- *Astigmatismus*: Die Brennweite ist für Punkte außerhalb der optischen Achse unterschiedlich groß.
- *Bildwölbung*: Die Abbildung auf der Bildebene ist gewölbt.

Weitere Fehler entstehen im Zusammenspiel der Linsen im Objektiv (z. B. Abweichungen von der optischen Achse). Die Fehler eines Objektivs lassen sich durch zwei typische Arten von Verzerrungen beschreiben (Abbildung 2.2) [Red00]. Viele Untersuchungen zeigten,

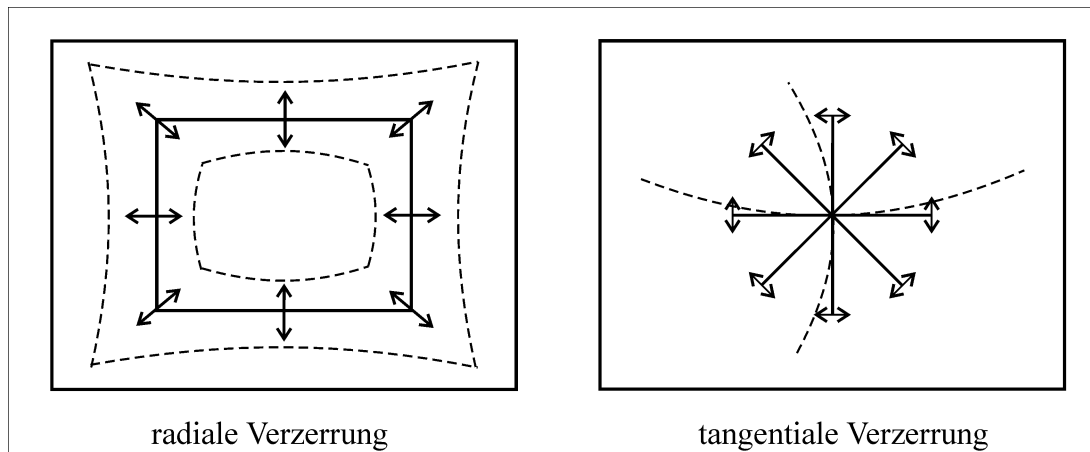


Abbildung 2.2: Radiale und tangentielle Verzerrungen der Linsenoptik

dass die tangentialen Verzerrungen gegenüber den radialen Verzerrungen vernachlässigbar sind [Red00]. Für eine allgemeine mathematische Modellierung dieser Verzerrungen eignet sich eine Betrachtung der Ebenen $z = -1$ und $z = 1$ im Koordinatensystem des Kameramodells (siehe Abbildung 2.3). In diesem Modell gilt für eine Lochkamera der einfache Zusammenhang:

$$\begin{bmatrix} x_D \\ y_D \\ z_D \end{bmatrix} = - \begin{bmatrix} x_U \\ y_U \\ z_U \end{bmatrix} \quad (2.1)$$

Tangentiale Verzerrungen können mittels eines Faktors Q für die x - und die y -Komponente beschrieben werden:

$$\begin{bmatrix} x_D \\ y_D \\ z_D \end{bmatrix} = - \begin{bmatrix} x_U \cdot Q \\ y_U \cdot Q \\ z_U \end{bmatrix} \quad (2.2)$$

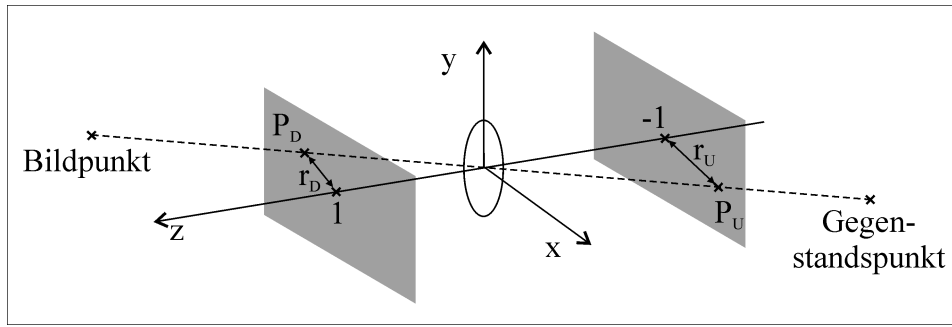


Abbildung 2.3: Kameramodell

Q ist gleichzeitig das Verhältnis der Radien r_D und r_U und kann mittels einer Potenzreihe mit geradzahligem Exponenten angegeben werden:

$$Q = \frac{r_D}{r_U}$$

$$r_D = r_u + K_3 \cdot r_U^3 + K_5 \cdot r_U^5 + \dots$$

$$Q = 1 + K_3 \cdot r_U^2 + K_5 \cdot r_U^4 + \dots \quad (2.3)$$

Für eine typische Verzerrung von fünf Bildpunkten bei einer Auflösung von 1000×1000 Pixel sind K_3 und K_5 ca. 1 für kleine Öffnungswinkel und ca. 0,1 für große Öffnungswinkel [Red00].

Modulationsübertragungsfunktion

Ein weiterer qualitätsbestimmender Parameter einer Kameraoptik resultiert aus deren begrenztem Auflösungsvermögen. Dieses wird durch die Modulationsübertragungsfunktion (*modulation transfer function*, MTF) beschrieben [Wen88]. Sie ist abhängig von:

- Lage des Bildpunktes
- Abbildungsmaßstab
- Blende
- Brennweite
- Wellenlänge des Lichtes

In Abbildung 2.4 ist die MTF in Bildmitte eines Fernsehkameraobjektives für zwei Brennweiten (a: $f = 100 \text{ mm}$ und b: $f = 40 \text{ mm}$) dargestellt.

2.2.2 Temporale und lokale Auflösung

Beginnend mit der Bildwandlung erfolgt eine lokale und temporale Diskretisierung des 2D-Bewegtbildes. Diese Diskretisierung wird durch die Parameter der lokalen und temporalen Auflösung repräsentiert. Die einzelnen Verarbeitungsschritte von konventionellem Video können mit unterschiedlich hohen Auflösungen arbeiten. Es ist jedoch nur die jeweils geringste Auflösung in der Verarbeitungskette für das Ende der Kette (Wiedergabe beim Nutzer) entscheidend.

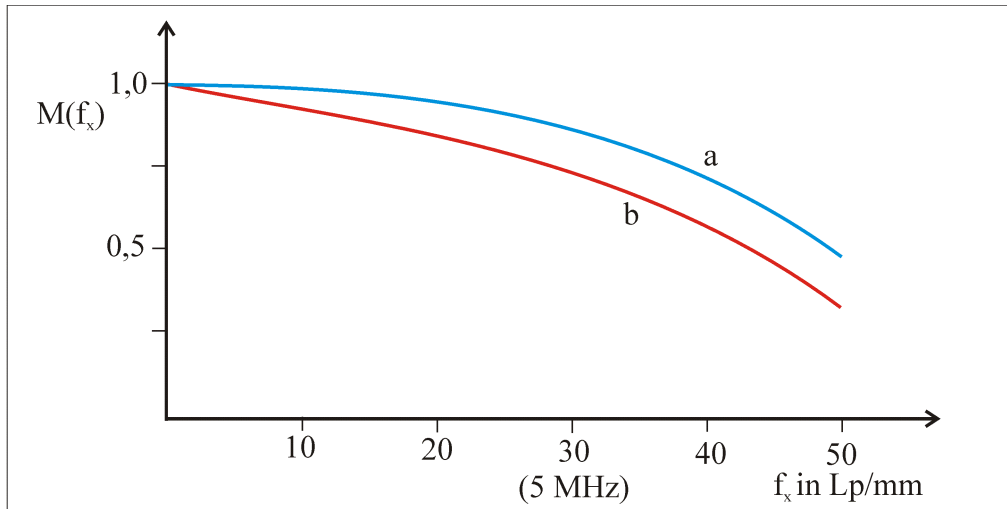


Abbildung 2.4: MTF eines Fernsehkameraobjektives (nach [Wen88])

Der Bildwandler einer Kamera tastet das Bildsignal $x(t)$ temporal diskretisiert in Intervallen mit der Zeitdauer T ab. Für die verschiedenen Bildwandlertypen geschieht dies nach unterschiedlichen Prinzipien. Der Elektronenstrahl einer Bildwandlerröhre tastet das Bild zeilenweise ab, während CCD-Matrizen elementweise ausgelesen werden. Dadurch bedingt werden entweder alle Bildpunkte zum selben Zeitpunkt t_0 oder zu unterschiedlichen Zeitpunkten innerhalb des Intervalls T aufgenommen. Die Belichtung erfolgt über die Dauer T eines Einzelbildes hinweg. Eine Ausnahme bilden Frametransfer-CCD-Bildwandler, da sie eine mechanische Shutterblende erfordern. Die temporale Abtastung wird durch die Bildwechselfrequenz $\omega = \frac{2\pi}{T}$ charakterisiert. Werden die Bildpunkte eines Einzelbildes gleichzeitig oder kontinuierlich nacheinander abgetastet, wird dies als *progressive* Abtastung bezeichnet. Für das Fernsehen wurde die *Zeilensprungabtastung* als ein Kompromiss aus Bandbreite und erforderlicher Bildwechselfrequenz eingeführt. Die Flimmerfusionsfrequenz beträgt bei hohen Adaptionsleuchtdichten bis zu 50 Hz [Hau94]. Daher muss zur Vermeidung von Großflächenflimmern die Bildwechselfrequenz mindestens in dieser Größenordnung liegen. Beim Zeilensprungverfahren werden z. B. zunächst alle ungeradzahigen Zeilen und anschließend alle geradzahigen Zeilen abgetastet und ebenso wiedergegeben. Dies verdoppelt die Bildwechselfrequenz bei gleicher Auflösung bzw. Bandbreite.

Bei der Wiedergabe wird die temporale Auflösung bei einigen Bildwandlertypen durch deren Nachleuchteffekte verringert. Beispielsweise weisen Plasmadisplays Nachleuchteffekte auf, welche Bewegungen im Bild als nicht kontinuierlich erscheinen lassen.

Die lokale Auflösung wird limitiert durch die lokale Abtastung des zweidimensionalen Bildsignals $x(r, s)$. Diese zweidimensionale Abtastung kann mit den Methoden und Algorithmen der *zweidimensionalen Signalverarbeitung* beschrieben werden [Hau94], [Ohm95], [Wan02], [Wen88] u. a. Wird das Bewegtbild $x(r, s, t)$ betrachtet, ist die Signalverarbeitung dreidimensional. Die zweidimensionale lokale Abtastung kann mit unterschiedlichen Rastern erfolgen. Typische homogene Raster sind rechteckig, hexagonal oder Quincunxförmig [Ohm95]:

$$\mathbf{D}_R = \begin{bmatrix} S/R & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{D}_H = \begin{bmatrix} 1 & 1/2 \\ 0 & \sqrt{3}/2 \end{bmatrix} \quad \mathbf{D}_Q = \begin{bmatrix} 1 & 1/2 \\ 0 & 1/2 \end{bmatrix} \quad (2.4)$$

mit R: horizontale Rasterweite, S: vertikale Rasterweite

In Bild 2.5 ist ein (rechteckiges) 3D-Abtastschema dargestellt. Es sind die horizontale und vertikale Rasterweite R und S sowie die Abtastdauer T gekennzeichnet.

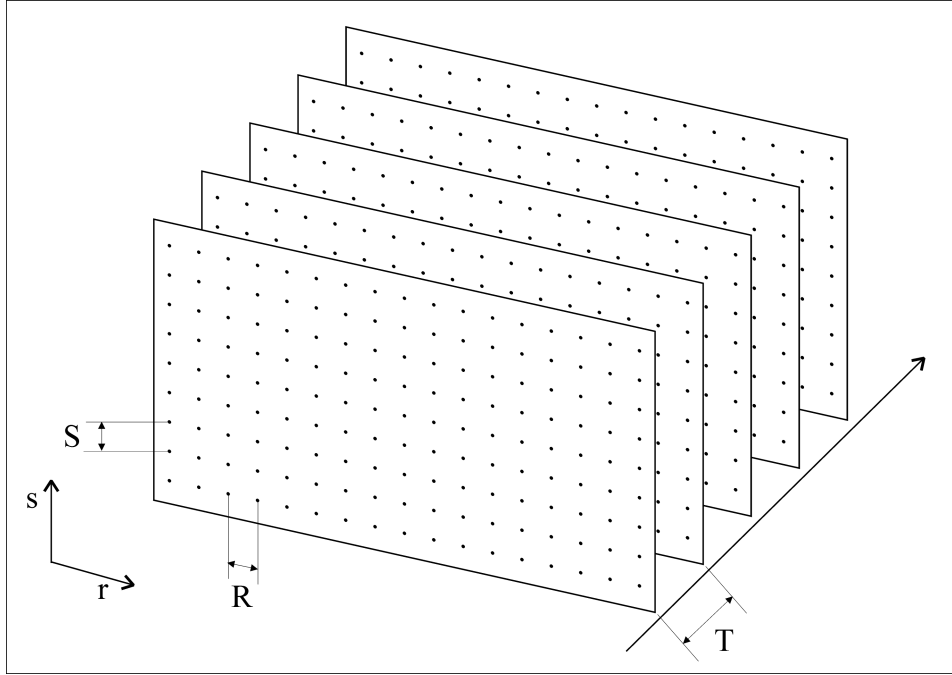


Abbildung 2.5: 3D-Abtastschema

Während die Zeilensprungabtastung einem Quincunx-Raster entspricht, sind in der digitalen Videosignalverarbeitung rechteckige Abtastraster üblich. Die rechteckige Abtastung entspricht der Multiplikation von $x(r, s)$ mit zwei Deltakämmen $\delta_R(r)$ und $\delta_S(s)$. Im Fourier-Spektralbereich ist dies eine 2D-Faltung $**$ und führt zu einem periodischen Spektrum [Ohm95]:

$$X(j\Omega_1, j\Omega_2) = X(j\omega_1, j\omega_2) ** \delta_{\omega_R, \omega_S}(\omega_1, \omega_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(m, n) \cdot e^{-jm \cdot \Omega_1} \cdot e^{-jn \cdot \Omega_2} \quad (2.5)$$

Um Aliaseffekte zwischen dem Signal $X(j\omega_1, j\omega_2)$ und dessen periodischer Fortsetzung¹ $X(j\Omega_1, j\Omega_2)$ zu vermeiden, ist das Originalsignal $x(r, s)$ in seiner Bandbreite dem Abtasttheorem gemäß zu begrenzen. Dies geschieht in Video- und Fernsehkameras mit Hilfe von optischen Tiefpässen, die aus doppelbrechenden Quarzscheiben bestehen. Das Resultat einer rechteckigen Abtastung ist eine rechteckige Bildpunktmatrix (M, N) . Dieses Parameterpaar oder auch dessen Produkt wird als (örtliche) *Auflösung* bezeichnet, z. B. (720, 576) gemäß ITU-R BT.601. Die lokale Auflösung kann ebenso im Frequenzbereich angegeben werden, z. B. mittels der maximalen horizontalen oder vertikalen Ortsfrequenz $f_{x,max}$, $f_{y,max}$. In praxi wird oft ausgehend von der maximalen Anzahl an abwechselnd schwarz-weißen Bildpunkten/-linien die Anzahl der Perioden pro Bildhöhe bzw. Bildbreite (jeweils halbe Linien-/Punktzahl) angegeben. Die Angabe erfolgt dann in *c/ph* (*cycles per picture height*).

Die genannten Maße für die Auflösung beziehen sich auf die Repräsentation des Bildes (z. B. der Videodatenstrom). Für ein wiedergabebezogenes Maß der Auflösung müssen die

¹mit $\omega_R = \frac{2\pi}{R}$, $\omega_S = \frac{2\pi}{S}$, $\Omega_1 = 2\pi \cdot \frac{\omega_1}{\omega_R}$ und $\Omega_2 = 2\pi \cdot \frac{\omega_2}{\omega_S}$

Bilddarstellung (Fernsehbildschirm, Kinoprojektion, etc.) und der Standort des Betrachters einbezogen werden. Nach Vorgabe eines typischen Betrachterabstandes (z. B. relativ – als Faktor der Bilddiagonale oder -höhe) kann die Auflösung in Grad angegeben werden (siehe Abbildung 2.6). Der typische Betrachterabstand von Fernsehbildschirmen beträgt das 2,0- bis 2,5-fache der Bildhöhe [Mäu95], [Bon96]. Für HDTV-Bildschirme wurde in [Bon96] das zweifache der Bildhöhe angenommen, sodass:

$$1 \text{ pel} \hat{=} 0,025 \text{ deg} \quad (2.6)$$

Diese betrachterabstandsabhängige Angabe kann ebenso auf eine Frequenz bezogen werden [Wan02]. Die Angabe erfolgt oft in *cpd* (*cycles per degree*).

$$f_{\theta} = \frac{f_y}{\theta} \quad (2.7)$$

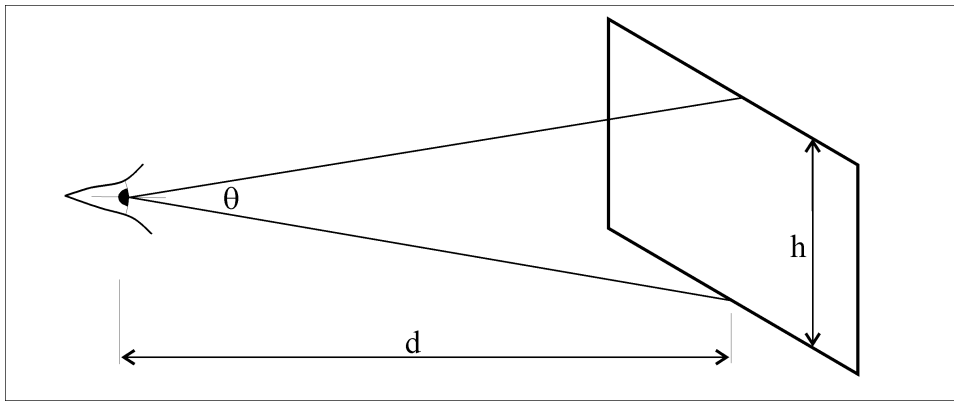


Abbildung 2.6: Bildbetrachtung: Betrachterabstand, Bildhöhe, Öffnungswinkel

Der visuelle Übertragungskanal weist für die Chrominanz eine 4...5-fach niedrigere Grenzfrequenz als für die Luminanz auf [Hau94]. Dies wird ausgenutzt, indem die Auflösung für die Chrominanzinformation in vielen Anwendungen geringer als die Auflösung für die Luminanz gewählt wird, z. B. (360, 576) für C_R , C_B nach ITU-R BT.601.

2.2.3 Luminanz

Das Luminanzsignal repräsentiert den Helligkeitseindruck des Visus. Die Qualität dieser Repräsentation wird durch die Amplitudenauflösung und die Gradation bestimmt. Für die Bildwandlung ist das maximal aufzeichenbare bzw. darstellbare Kontrastverhältnis ein weiterer Qualitätsparameter.

Amplitudenauflösung

Die Diskretisierung der Amplituden führt zu einer begrenzten Amplitudenauflösung. Der menschliche Visus kann Luminanzunterschiede von bis zu 1% über einen großen Bereich der Luminanz unterscheiden [Hau94]. Der wahrnehmbare Kontrastumfang beträgt etwa 100 : 1. Damit ergibt sich die erforderliche Anzahl an linearen Quantisierungsstufen zu: $\lg 100 / \lg(1 + 1\%) = 462,82 < 463$. Durch eine exponentielle Verzerrung der Kennlinie (γ -Korrektur, s. u.) sind für die Quantisierung der Luminanz jedoch $2^8 = 256$ Stufen ausreichend. Daher sehen die meisten Standards eine Quantisierungsbreite B von 8 Bit

vor. Soll eine Weiterbearbeitung erfolgen, bei der eine Kaskadierung von Rundungsfehlern zu erwarten ist, wird mitunter eine Quantisierungsbreite B von 10 Bit verwendet. Es stehen jedoch nicht grundsätzlich alle 2^B Quantisierungsstufen zur Verfügung, da für viele Anwendungen im oberen und unteren Signalbereich mehrere Quantisierungsstufen reserviert sind (z. B. nach ITU-R BT.601 für Luminanz 220 Quantisierungsstufen).

γ -Korrektur (Gradation)

Die Bildwandlungen auf Aufnahme- und Wiedergabeseite erfolgen auf Grund der physikalischen Eigenschaften oft nichtlinear. Der Zusammenhang zwischen der Eingangsspannung U und der Leuchtdichte L_{CRT} einer Kathodenstrahlröhre kann angenähert werden durch:

$$L_{CRT} = U^{\gamma_{CRT}} \quad (2.8)$$

Der Wert für γ_{CRT} ist typisch 2,2...2,5. Auch der Zusammenhang zwischen der Belichtung H und der Ausgangsspannung U_C bei einer Bildaufnahmeröhre ist nichtlinear.

$$U_C = H^{\gamma_C} \quad (2.9)$$

Der Wert für γ_C beträgt etwa 1,7. Für CCD-Wandler ist der Zusammenhang nahezu linear [Wan02].

Für die Fernsehübertragung wurde vereinbart, die so genannte γ -Korrektur auf der Senderseite durchzuführen, sodass bereits die Repräsentation diese Charakteristik aufweist. Eine fehlerhafte γ -Korrektur führt zu falschen Kontrastverhältnissen. Da das Kontrastempfinden individuell unterschiedlich ist und von weiteren Größen abhängig ist (Umgebungsbeleuchtung, Adaption u. a.), kann eine Qualitätsbeeinflussung quantitativ nur beschränkt erfasst werden.

Maximales Kontrastverhältnis

Ein weiterer Parameter ist das maximale Kontrastverhältnis, welches der Bildwandler wandeln kann. Der wahrnehmbare Kontrastumfang beträgt mindestens 100 : 1. Die Technologie der Bildaufnahmewandler (CCD) ist fortgeschritten, sodass ein Umfang von $> 2^{8...9}$ aufgezeichnet werden kann. Auf der Bildwiedergabeseite wird der wahrnehmbare Umfang jedoch nicht immer erreicht². Dies wird insbesondere im direkten Vergleich als Mangel empfunden.

2.2.4 Farbrepräsentation

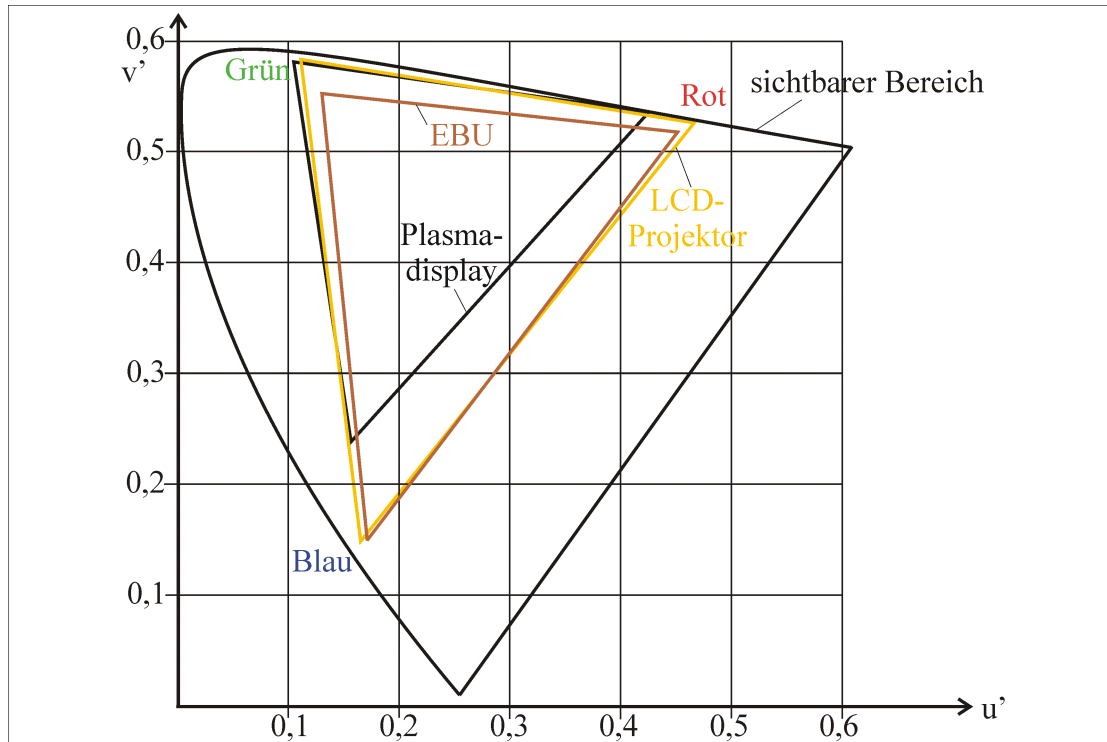
Die Aufnahme, Verarbeitung und Wiedergabe von Bewegtbild erfolgt meist nicht für den gesamten Farbraum, welchen der Visus wahrnehmen kann. Die Ursache dafür ist, dass die Bildwandler auf Grund ihres technischen Aufbaues nur einen Teil des sichtbaren Farbraumes wandeln können. Dieser wird limitiert durch die verwendeten Grundfarben (meist Rot, Grün, Blau) des Farbraumes. Für die Angabe der Primärvalenzen wird oft der xyz-Farbraum der CIE genutzt, mittels dessen alle sichtbaren Farben als additive Mischung beschrieben werden können. Ausgehend von den Normfarbwerten X, Y, Z ergeben sich $x = X/(X + Y + Z)$, $y = Y/(X + Y + Z)$ und $z = Z/(X + Y + Z)$. In Tabelle 2.1 sind die Grundfarben Rot, Grün und Blau entsprechend der EBU [EBU75] und der FCC angegeben.

²z. B. für CRT 40 : 1 bis 200 : 1, ältere Plasmadisplays 50 : 1

	EBU			FCC		
	Rot	Grün	Blau	Rot	Grün	Blau
x	0,64	0,29	0,15	0,67	0,21	0,14
y	0,33	0,60	0,06	0,33	0,71	0,08
z	0,03	0,11	0,79	0,00	0,08	0,78

Tabelle 2.1: Festlegungen der Primärvalenzen nach EBU und FCC

Zur Darstellung des repräsentierbaren Farbraumes werden häufig die x, y - und die u', v' -Farbtafeln gemäß CIE verwendet. Die u', v' -Farbtafel wurde zur Anpassung der euklidischen Abstände definiert mit: $u' = 4 \cdot X / (X + 15 \cdot Y + 3 \cdot Z)$ und $v' = 9 \cdot X / (X + 15 \cdot Y + 3 \cdot Z)$. In Abbildung 2.7 ist eine u', v' -Farbtafel dargestellt. Das so genannte Hufeisen umfasst den sichtbaren Bereich. Es wird deutlich, dass die typischen RGB-Systeme (EBU) einen Großteil der sichtbaren Farben nicht darstellen können. Zum Vergleich sind die darstellbaren Farbräume eines Plasmadisplays und eines LCD-Projektors eingezeichnet.

Abbildung 2.7: CIE- u', v' -Farbtafel

2.2.5 Bandbreite und Übertragungsfehler

Die Qualität von Bewegtbild wird ebenso durch die zur Verfügung stehende Bandbreite und Übertragungsfehler eingeschränkt. Dies galt bereits für die analoge Übertragung. Parameter wie die Frequenzbandbreite und S/N beeinflussen Qualitätsmerkmale, wie z. B. Bildrauschen, FM-„Fischchen“, SECAM-„Feuer“.

Die Qualität einer digitalen Videocodierung und -übertragung wird ebenfalls durch die zur Verfügung stehende Datenrate und auftretende Bitfehler eingeschränkt. Die damit im Zusammenhang stehenden Qualitätsparameter hängen von der verwendeten Codierung

ab. Da die z. Z. etablierten Videocodierverfahren (MPEG-2, ITU H.263) auf der Diskreten Cosinus-Transformation basieren, sind die auftretenden Störungstypen meist ähnlich. Nach [ANSI96a] sind dies u. a. Blockfehler, Unruhe an Kanten, Rucken, Bewegungsartefakte.

2.3 Subjektive Bewertung von 2D-Bewegtbild

Die Aufzeichnung, Übertragung und Wiedergabe von Bewegtbild für das Fernsehen oder für eine Multimedia-Anwendung zielt selbstverständlich auf die Nutzung durch den Menschen ab. Daher ist die Bewertung der Qualität des Bewegtbildes durch den Menschen maßgebend. Mit der technischen Entwicklung des Fernsehens wuchs der Bedarf an Methoden für die subjektive Qualitätsbeurteilung. In den Fernsehanstalten und den Laboratorien wurden diverse Verfahren entwickelt, die zwar in verschiedenen Aspekten übereinstimmen, jedoch noch keine austauschbaren Ergebnisse lieferten. Dies machte eine Standardisierung erforderlich. Die Empfehlung ITU-R BT.500 [ITU02] findet vielfach Anwendung. Sie wurde bereits im Jahre 1974 erstmalig verabschiedet und liegt seit 2002 in der Version 11 vor. Sie beschreibt eine Methodik zur subjektiven Qualitätsbewertung des Fernsehbildes, welche auf vielfach etablierten Verfahren zur subjektiven Qualitätsbewertung beruht. Video wird zunehmend auch in Kommunikations- und Multimedia-Anwendungen genutzt. Geeignete subjektive Testverfahren werden in den Empfehlungen ITU-T P.910 [ITU99], ITU-T P.920 [ITU00] u. a. vorgeschlagen.

2.3.1 Anzahl und Auswahl der Betrachter

Eine Vielzahl von Untersuchungen und Erfahrungen führte zu dem Konsens, dass für die subjektive Bewertung von Bewegtbild mindestens etwa 10...20 Betrachter notwendig sind. Diese Aussage beruht auf der Beobachtung, dass die Verhältnisse der geometrischen Mittel der Bewertungen sich ab etwa 15 Versuchspersonen stabilisieren [Sch94]. Diese Anzahl ist für die repräsentative Nachbildung der Zielgruppe ausreichend. Die relevanten ITU-Empfehlungen spiegeln diese Erkenntnis wider. Nach ITU-R BT.500 [ITU02] sind zur subjektiven Bewertung von Fernsehbild mindestens 15 Betrachter notwendig. Die Bewertung von Bewegtbild in Multimedia-Anwendungen ist nach ITU-T P.910 mit 4...40 Betrachtern möglich, während im Allgemeinen mindestens 15 Betrachter notwendig sind. Weiterführende Untersuchungen bestätigten diese Forderung [Fec00].

Die Testpersonen sollen einen normalen oder einen auf normal korrigierten Visus besitzen. Die Eignung des Visus kann durch eine Überprüfung der Sehschärfe und der Farbtauglichkeit nachgewiesen werden [ITU02], [ITU99], [Sch94]. Hierfür kommen die in der Ophthalmologie üblichen Testtafeln³ zum Einsatz.

Des Weiteren sollen die Testpersonen keine Experten in der Bewertung von Bewegtbild sein. Dies läge beispielsweise vor, wenn sie sich beruflich mit der Bildbewertung beschäftigen oder wenn sie Erfahrungen in der Bildbewertung haben [ITU02], [ITU99]. Die Gültigkeit der groben Einteilung in Laien und Experten wurde mehrfach nachgewiesen [Sch94]. Experten können kleinere Qualitätsunterschiede differenzieren und die Streuung der Ergebnisse ist deutlich kleiner als bei Laien. Dies hat dazu geführt, dass in praxi

³z. B. Landoltringe zur Messung der Sehschärfe und Testtafeln nach Ishihura zur Ermittlung der Farbtauglichkeit

oft Tests mit Experten durchgeführt werden, wohingegen die eigentliche Zielgruppe natürlich die Laien sind. Jedoch können mit Hilfe von wenigen Experten im Vorfeld die Resultate eines Tests bereits prognostiziert werden. Eine feinere Unterteilung der Bewertungsergebnisse anhand von Gruppen (z. B. nach Alter, Geschlecht, Beruf) ist bei einer entsprechend großen Anzahl von Testpersonen möglich. Die ITU-R BT.500 empfiehlt zumindest die Aufnahme von charakterisierenden Daten wie Beruf (z. B. Tätigkeit in einer Fernsehanstalt, Student, Verwaltungsangestellter), Geschlecht und Alter.

2.3.2 Testbedingungen

Um vergleichbare Ergebnisse der subjektiven Tests zu erhalten, müssen die Randbedingungen definiert sein. Dies betrifft neben den technischen Parametern der Wiedergabegeräte und deren Anordnung auch die Eigenschaften der Umgebung.

Die technischen Parameter der Wiedergabegeräte sind entsprechend den Empfehlungen konkret für jede Testreihe anzupassen und während des Verlaufes einzuhalten [ITU99]. Für die Bewertung von Fernsehbildern unterscheidet die Empfehlung ITU-R BT.500-11 zwischen Laborbedingungen und heimischen Bedingungen.

In der Empfehlung ITU-T P.910 werden eine direkte und eine indirekte Methode zur Bewertung von Video in Multimedia-Anwendungen beschrieben. Bei der indirekten Methode werden die wiedergegebenen Testsequenzen mit einem hochwertigen Aufzeichnungssystem aufgenommen, um anschließend die Aufzeichnungen zur Bewertung zu benutzen.

Betrachtungsabstand, -winkel

Die PVD (*preferred viewing distance*) nach [ITU02] als Verhältnis des Abstandes zur Bildhöhe nimmt mit abnehmender Bildhöhe zu (siehe Abbildung 2.8). Sie beträgt das 3...9-fache der Bildhöhe. Da Multimedia-Anwendungen auf unterschiedlichere Weise genutzt werden, wird nach [ITU02] ein Betrachtungsabstand im Bereich der 1...8-fachen Bildhöhe empfohlen, der entsprechend der konkreten Anwendung zu spezifizieren ist. Der

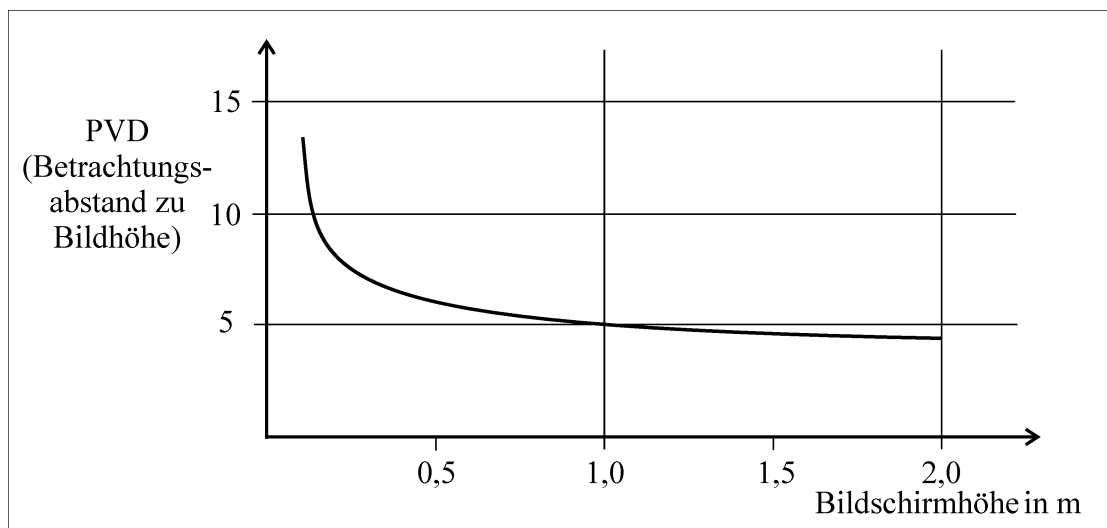


Abbildung 2.8: Betrachtungsabstand zur Bewertung von Fernsehbildern nach [ITU02]

Betrachtungswinkel beeinflusst den Bildeindruck und muss daher limitiert werden. Während er für Kathodenstrahlröhren auf z. B. 30° begrenzt ist, kann für neue Bildschirmtechnologien (z. B. LCD) noch keine einheitliche Empfehlung gegeben werden.

Bildschirmhelligkeit, Umgebungsbeleuchtung

Um in allen Tests jeweils den gleichen Kontrasteindruck beim Betrachter durch das Testmaterial erzeugen zu können, müssen die Helligkeit des Bildschirmes und der Umgebung definiert sein. Die Leuchtdichte hinter dem Bildschirm darf maximal das 0,2-fache des Spitzenwertes auf dem Bildschirm betragen [ITU99]. Die Farbtemperatur des Hintergrundes sollte der des Bildschirmes entsprechen (meist tageslichtweiß). Die restliche Umgebungsbeleuchtung sollte gering sein (≤ 20 lx). Die Leuchtdichte der inaktiven Anteile des Bildschirmes dürfen maximal das 0,02- bzw. 0,05-fache des Spitzenwertes betragen.

2.3.3 Auswahl des Testmaterials

Die Auswahl des Testmaterials hat so zu erfolgen, dass die zu ermittelnden Kriterien in ihrem Wertebereich erfasst werden. Für die allgemeine Qualitätsbewertung wird „kritisches, aber nicht übermäßig kritisches“ Testmaterial empfohlen [ITU02]. Sollen hingegen bestimmte Schwächen eines Systems ermittelt werden, muss das Testmaterial hinsichtlich der Kriterien variieren. Für die Entwicklung eines neuen Bildqualitätsmodells wird jedoch Testmaterial benötigt, welches den gesamten Bereich von einfachen bis hin zu komplexen Bildern abdeckt [Fec00].

Für viele Tests wird ein Satz an Testsequenzen verwendet, der ursprünglich gemäß der Empfehlung ITU-R BT.802-1 [ITU94b] und nun entsprechend der SMPTE-Praxisempfehlung RP 222 [SMPTE03] angeboten wird. Insbesondere bei der Entwicklung der Videokompressionsverfahren (MPEG) wurden diese Sequenzen⁴ vielfach verwendet, da hier die typischen Kompressionsartefakte auftreten.

Für die Entwicklung eines allgemeingültigen Modells zur objektiven Videoqualitätsbeurteilung stellte die VQEG einen umfangreichen Satz an Testsequenzen zusammen, der einen großen Qualitätsbereich umfasst [VQEG03].

2.3.4 Single-Stimulus-Methoden

Bei den Single-Stimulus-Methoden wird den Betrachtern jeweils nur eine Bildsequenz gezeigt. Der Betrachter hat die Sequenz auf einer Bewertungsskala unabhängig zu bewerten. Diese Methode findet Verwendung, wenn eine bestimmte Eigenschaft (z. B. Schärfe) oder Störung (z. B. Blockstörung) entsprechend ihrer Güte bzw. Sichtbarkeit bewertet werden soll. In den ITU-Empfehlungen ITU-R BT.500-11 und ITU-T P.910 werden die Single-Stimulus Methoden SS, SSMR, SSCQE und ACR beschrieben.

SS (Single-Stimulus)

Die Präsentation der Testsequenz erfolgt bei dieser Methode gemäß [ITU02] genau einmal. Zwischen den Testsequenzen (z. B. 10 Sekunden) wird eine 50%-Grau-Fläche gezeigt. Bei dieser Methode ist die Bewertung abhängig von der gewählten Reihenfolge der Testsequenzen.

SSMR (Single Stimulus with Multiple Repetition)

Um von der Reihenfolge der Testsequenzen unabhängige Bewertungen erwarten zu können, wird bei der abgeleiteten SSMR-Methode die Testreihe zweimal wiederholt, wobei die einzelnen Testsequenzen in unterschiedlichen Reihenfolgen gezeigt werden. Für diese beiden Methoden werden vier Bewertungstypen beschrieben. Während Typ 1 eine fünfstufige

⁴z. B. „Mobile and Calendar“, „Flower Garden“, „Susie“

verbale Bewertungsskala für die Bildgüte bzw. Bildstörungen vorsieht (siehe Tabelle 2.2) wird bei Typ 2 eine elfstufige numerische Skala verwendet. Typ 3 beschreibt den Verzicht auf jegliche Kategorie für spezielle Anwendungen und Typ 4 die Bewertung auf Grund von bestimmten Leistungsfähigkeiten (Lesbarkeit eines Textes, Erkennbarkeit eines Objektes etc.).

ACR (Absolute Category Rating)

Die ACR-Methode nach [ITU99] dient zur Bewertung von Video in Multimedia-Anwendungen. Sie ähnelt im Ablauf der SS-Methode und verwendet dieselbe Skala zur Bewertung der Bildgüte (Tabelle 2.2).

Quality	Impairment
5 Excellent	5 Imperceptible
4 Good	4 Perceptible, but not annoying
3 Fair	3 Slightly annoying
2 Poor	2 Annoying
1 Bad	1 Very annoying

Tabelle 2.2: Bewertungsskala nach [ITU02]

SSCQE (Single Stimulus Continuous Quality Evaluation)

Die SSCQE-Methode nach [ITU02] wurde für die Bewertung von digitalen Bildsequenzen entwickelt. Diese weisen auf Grund ihrer Codierung szenenabhängige und zeitveränderliche Störungen auf, sodass eine kontinuierliche Bewertung notwendig wird. Bei der SSCQE-Methode bedient der Betrachter einen Schiebester, dessen Skala in die fünf Abschnitte nach Tabelle 2.2 unterteilt ist. Der Schiebester muss einen Stellbereich von 10 cm aufweisen und darf nicht zurückschnellen. Durch die Bedienung des Schiebesters während der Testsequenz kann der Betrachter eine kontinuierliche Bewertung vornehmen. Diese wird zweimal pro Sekunde erfasst.

2.3.5 Double-Stimulus-Methoden

Eine weitere Klasse von Methoden zur subjektiven Qualitätsbewertung sieht einen Vergleich mit einer zweiten Bildsequenz vor. Diese Methoden eignen sich beispielsweise zur Bewertung eines Übertragungskanal. In den ITU-Empfehlungen ITU-R BT.500-11 und ITU-T P.910 werden die Double-Stimulus Methoden DSIS (Double-Stimulus Impairment Scale), DSCQS (Double-Stimulus Continuous Quality-Scale), Stimulus-Comparison, SDSCE (Simultaneous Double Stimulus for Continuous Evaluation), DCR (Degradation Category Rating) und PC (Pair Comparison) vorgeschlagen.

DSIS (Double-Stimulus Impairment Scale), DCR (Degradation Category Rating)

Die DSIS-Methode der EBU nach [ITU02] ist eine der wichtigsten Methoden zur Bewertung der Signalverarbeitung/-übertragung. Der Betrachter bekommt immer zunächst die Referenz und anschließend die zu bewertende Sequenz für jeweils 10 s gezeigt. Zwischendurch wird für 3 s eine ca. 20%-Grau-Fläche gezeigt. Alternativ können beide Sequenzen doppelt gezeigt werden. Der Betrachter hat die Störung auf einer fünfstufigen Skala gemäß Tabelle 2.2 zu bewerten. Da die DSIS-Methode etabliert ist, wurde sie als DCR-Methode nach [ITU99] zur Bewertung der Videoqualität in Multimedia-Anwendungen standardisiert.

DSCQS (Double-Stimulus Continuous Quality-Scale)

Bei der DSCQS-Methode wird im Gegensatz zur DSIS-Methode die Referenz nicht grundsätzlich als erste Sequenz gezeigt, sondern die Auswahl von Referenz und zu bewertender Sequenz erfolgt ohne Kenntnis des Betrachters. Er hat beide Sequenzen jeweils paarweise zu bewerten (siehe Abbildung 2.9). Die DSCQS-Methode ist besonders empfehlenswert, wenn die zu bewertenden Störungen nur geringfügige Unterschiede erzeugen bzw. sich durch die Simulation der Störungen nur geringfügige Unterschiede erzeugen lassen.

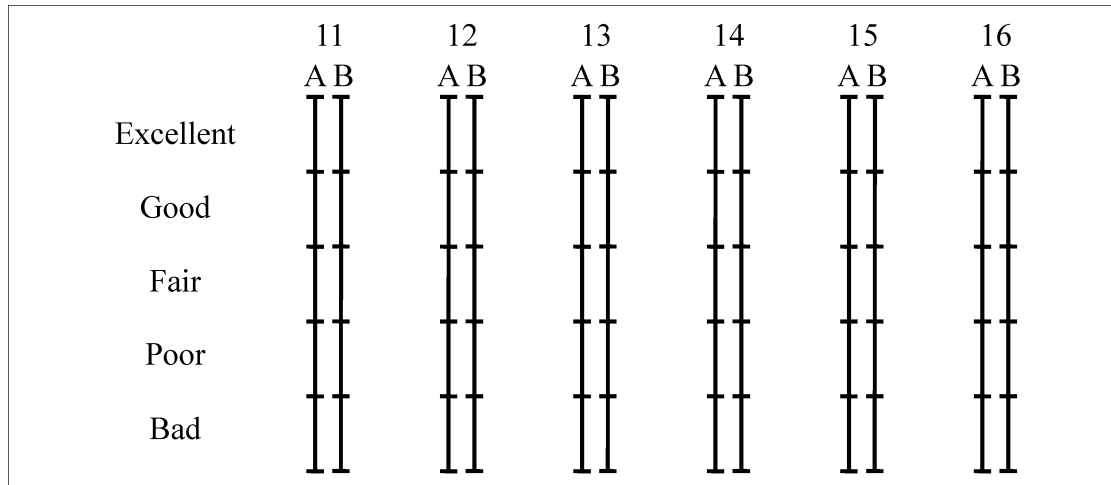


Abbildung 2.9: DSCQS-Skala

Stimulus-Comparison

Die Stimulus-Comparison-Methode wird zum direkten Vergleich von zwei Testsequenzen verwendet. Die Sequenzen können entweder auf einem oder auf zwei Monitoren gezeigt werden. Der Ablauf gleicht dem der Single-Stimulus-Methoden. Es werden drei Varianten der Bewertung unterschieden:

1. Der Betrachter hat zu entscheiden, in welcher Relation die Qualitäten der Sequenzen zueinander stehen. Hierzu ist folgende Bewertungsskala zu verwenden:
 - 3 Much worse
 - 2 Worse
 - 1 Slightly worse
 - 0 The same
 - +1 Slightly better
 - +2 Better
 - +3 Much better
2. Der Betrachter hat die Relation der Qualitäten ohne eine Kategorienskala zu bewerten. Hierfür hat er für jede Relation einen Punkt auf einer freien Skala zu zeichnen oder eine Zahl anzugeben.
3. Wenn ein bestimmtes Merkmal bewertet werden soll, werden dem Betrachter zwei Sequenzen gezeigt, welche das Merkmal in einem unterschiedlichen Maße aufweisen. Auch die Dauer der Bewertung ist ein Maß für die Relation zwischen den Sequenzen.

SDSCE (Simultaneous Double Stimulus for Continuous Evaluation)

Die SDSCE-Methode ist eine Weiterentwicklung der SSCQE-Methode, die zur Bewertung von zeit- oder szenenabhängigen Fehlern geeignet ist. Wenn die Wiedergabetreue eines Systems hinsichtlich dieser Art von Fehlern untersucht wird, muss ein Vergleich zu einer Referenz erfolgen. Hierfür werden bei der SDSCE-Methode die Referenz- und die Testsequenz gleichzeitig auf einem oder auf zwei Monitoren gezeigt. Die Bewerber bedienen einen Schieberegler entsprechend der SSCQE-Methode, dessen Anschläge der maximalen bzw. der minimalen Wiedergabetreue entsprechen.

PC (Pair Comparison)

Die PC-Methode nach [ITU99] dient zum paarweisen Vergleich von Testsequenzen in Multimedia-Anwendungen. Sollen mehr als zwei Testsequenzen bewertet werden, müssen die Vergleichspaare alle möglichen Kombinationen der Testsequenzen repräsentieren. Die Sequenzpaare werden ähnlich wie bei DSIS sequentiell präsentiert. Wenn die Sequenzen eine niedrige Auflösung haben, können sie dagegen zeitgleich auf nur einem Monitor gezeigt werden. Der Betrachter hat zu entscheiden, welche Testsequenz er bevorzugt.

2.3.6 Objektbasierte Bewertung

Mit der Einführung der objektbasierten Codierung audiovisueller Inhalte [ISO98] wurden auch Richtlinien zur objektbasierten Bewertung notwendig. In [ITU99] wird eine Methodik dafür empfohlen. Um sowohl die einzelnen Objekte als auch das zusammengesetzte Bild beurteilen zu können, werden zwei Testläufe vorgeschlagen. In einem ersten Testlauf werden die aus den Objekten zusammengesetzten Bilder beurteilt. Hierfür sind die eingeführten Verfahren wie ACR oder DCR anzuwenden. In einem zweiten Testlauf werden dieselben Objekte einzeln vor einem grauen Hintergrund bewertet (*object-based evaluation*, OBE). Für die Gesamtbewertung sollen die Ergebnisse der beiden Testläufe geeignet zusammengefasst werden.

Diese Zusammenfassung der Teilergebnisse wird in [ITU99] nicht erläutert. Die Findung eines guten Kompromisses muss entsprechend der jeweiligen Anwendung der Videoobjekte erfolgen. So ist beispielsweise für mehrfach zu verwendende Objekte die Qualität der Silhouette sehr wichtig, sodass die Einzelbewertung stärker zu wichten wäre.

2.3.7 Analyse von subjektiven Bewertungsergebnissen

Entsprechend dem Untersuchungsziel erfolgt eine Auswertung der Bewertungsergebnisse, für die folgende Teilschritte typisch sind:

- Ausschluss bzw. Normierung von Bewertern, die in ihrer Bewertung konstant von der durchschnittlichen Bewertung abweichen
- Ausschluss von Bewertern, die in ihrer Bewertung beliebig von der durchschnittlichen Bewertung abweichen
- Untersuchung, ob die Bewertungsergebnisse normalverteilt sind
- Bestimmung des Vertrauensintervalls der Ergebnisse
- Bestimmung eines analytischen Zusammenhanges zwischen der durchschnittlichen Bewertung und einem objektiven Bewertungsmaß

Angangspunkt jeder Analyse ist die mittlere Bewertung (*mean opinion score* - MOS) \bar{u}_{jkr} und deren Standardabweichung S_{jkr} :

$$\bar{u}_{jkr} = \frac{1}{N} \sum_{i=1}^N u_{ijkr} \quad (2.10)$$

u_{ijkr} - Bewertung von Bewerber i für Test j , Sequenz/Bild k , Wiederholung r
 N - Anzahl der Bewerber

$$S_{jkr} = \sqrt{\sum_{i=1}^N \frac{(\bar{u}_{jkr} - u_{ijkr})^2}{N-1}} \quad (2.11)$$

Die Bewertungen u_{ijkr} erfordern zunächst eine Normierung, da die Bewerber die ihnen zur Verfügung stehende Bewertungsskala unterschiedlich ausnutzen. Nach [ITU02] wird folgende Skalierung vorgeschlagen:

$$p = (\bar{u} - u_{min}) / (u_{max} - u_{min}) \quad (2.12)$$

Die Variablen u_{min} und u_{max} stehen hierbei für die schlechteste bzw. beste Bewertung. Diese Skalierung wurde u. a. in [VQEG03] zur Skalierung von DSCQS-Ergebnissen verwendet, wobei u_{max} der maximalen Differenz, d. h. der schlechtesten Bewertung entspricht. Eine Normierung ist insbesondere auch bei den Single-Stimulus-Methoden erforderlich, da den Bewertern keine Referenz zur Verfügung steht. Für SSCQE-Methoden wird nach [Fec00] das so genannte Z-Scoring verwendet, welches den Mittelwert und die Varianz aller Bewerber aufeinander abgleicht. Hierfür werden die Mittelwerte und Varianzen eines jeden Bewerbers bestimmt, um damit die Messwerte zu normieren. Die normierten Messwerte werden mit einer linearen Transformation auf die ursprüngliche Skala abgebildet.

Die Normalverteilung kann z. B. mittels der Kurtosis-Koeffizienten überprüft werden. In [ITU02] werden Methoden (für DSIS, DSCQS, SSCQE) beschrieben, mit der einzelne Bewertungen ausgeschlossen werden, wenn diese nicht normalverteilt sind.

Durch die Bestimmung des Vertrauensintervalls $[\bar{u}_{jkr} - \delta_{jkr}, \bar{u}_{jkr} + \delta_{jkr}]$ kann die Präzision der Messung eingeschätzt werden. Der wahre Wert liegt mit einer 95%igen Wahrscheinlichkeit in diesem Intervall. Je kleiner δ_{jkr} ist, desto genauer ist die Messung. Es ergibt sich aus der Standardabweichung wie folgt:

$$\delta_{jkr} = 1,96 \cdot \frac{S_{jkr}}{\sqrt{N}} \quad (2.13)$$

Oftmals sollen mit Hilfe subjektiver Tests Modelle für objektive Maße überprüft werden [ITU02], [Fec98], [VQEG03]. Hierfür eignet sich eine logistische Regression, wie sie auch zur Beschreibung von Wachstumsprozessen benutzt wird. Dazu ist eine Annäherung an die logistische Funktion $p = f(D)$ vorzunehmen:

$$p = \frac{1}{1 + e^{(D-D_M) \cdot G}} \quad (2.14)$$

Die Bestimmung der Konstanten G und D_M kann mittels der Logit-Transformation erfolgen:

$$\frac{1}{p} - 1 = e^{(D-D_M) \cdot G}$$

$$\begin{aligned} \ln\left(\frac{1}{p} - 1\right) &= (D - D_M) \cdot G \\ \frac{D - D_M}{\ln\left(\frac{1}{p} - 1\right)} &= \frac{1}{G} \end{aligned} \quad (2.15)$$

Damit ist das Problem auf eine lineare Regression reduziert.

Die oben genannte logistische Funktion ist symmetrisch. Sie ist geeignet, wenn das objektive Maß D in einem Verhältnismaß angegeben wird (z. B. dB). Für absolute Maße (z. B. ms) empfiehlt sich eine nichtsymmetrische Funktion [ITU02]:

$$p = \frac{1}{1 + \left(\frac{d}{d_m}\right)^{\frac{1}{G}}} \quad (2.16)$$

2.4 Objektive Maße zur Bewertung von 2D-Bewegtbild

Die im vorherigen Abschnitt beschriebenen Methoden der subjektiven Qualitätsbewertung von Bewegtbild sind in ihrer Durchführung sehr aufwändig. Daher besteht ein großes Interesse, eine Bewertung mittels objektiver Maße zu erzielen, welche der subjektiven Beurteilung möglichst nahe kommt. Objektive Maße haben den Vorteil, dass sie mühelos mit Hilfe von Messgeräten gemessen werden können. Sie sind deshalb oft schon während der Laufzeit bekannt und können so beispielsweise zur Steuerung des Encoders genutzt werden.

In den letzten Jahrzehnten wurde eine Vielfalt an objektiven Qualitätsmaßen entwickelt, die auf unterschiedlichen Ansätzen beruhen und mit unterschiedlichen Zielstellungen konzipiert wurden. In den folgenden Unterabschnitten wird eine relevante Auswahl an objektiven Qualitätsmaßen vorgestellt und abschließend eine Systematisierung vorgenommen.

2.4.1 Signaltechnische Maße

Im einfachsten Fall kann Bewegtbild nach signaltechnischen Kriterien bewertet werden. Diese werden häufig angewendet, da im Allgemeinen die Forderung besteht, dass ein objektives Qualitätsmaß mit vertretbarem Aufwand zu berechnen ist. Hierfür erfolgt meist ein bildpunktweiser Vergleich zwischen der gestörten Bildsequenz und der Referenzbildsequenz. Der mittlere quadratische Fehler (*mean squared error*, MSE) einer Bildsequenz k, m, n wird wie folgt ermittelt:

$$MSE = \frac{1}{N} \sum_{k,m,n} (\Psi_1(m, n, k) - \Psi_2(m, n, k))^2 \quad (2.17)$$

Als Qualitätsmaß wird das logarithmisierte Verhältnis der maximalen quadratischen Amplitude zum MSE genutzt (*peak signal-to-noise ratio*, PSNR in dB):

$$PSNR = 10 \cdot \lg \frac{\Psi_{max}^2}{MSE} \quad (2.18)$$

Mitunter wird zur Vereinfachung der Berechnung statt des MSE der absolute Betrag der Abweichung betrachtet (*mean absolute difference*, MAD):

$$MAD = \frac{1}{N} \sum_{k,m,n} |\Psi_1(m, n, k) - \Psi_2(m, n, k)| \quad (2.19)$$

Diese Maße korrelieren in keinem hohen Maße mit der subjektiven Bewertung (z. B. PSNR mit einem Korrelationsfaktor von nur ca. 0,6). Dennoch werden diese Maße seit vielen Jahren vielfach in der Bild- und Videoverarbeitung als objektive Qualitätsmaße genutzt [Kay96], [Wan02].

Der PSNR-Wert korreliert mit der subjektiven Wahrnehmung besonders wenig, wenn Fehler in Form eines Amplitudenoffsets oder eines Bildpunktversatzes auftreten. In Ta-

Störung	diagonaler Sprungverlauf	diagonale Sinusperiode (5 Perioden)	diagonale Sinusperiode (10 Perioden)
Schwarz-Niveau Verschiebung um 1,4%	37,8 dB	37,8 dB	37,8 dB
Verstärkung um 1,8%	38,9 dB	37,9 dB	37,8 dB
horizontaler Versatz um 0,5 Bildpunkte	48,1 dB	33,8 dB	27,8 dB
horizontaler Versatz um 1 Bildpunkt	48,1 dB	27,8 dB	21,8 dB

Tabelle 2.3: PSNR-Werte für einige Störungen

belle 2.3 sind die PSNR-Werte infolge eines bildeinheitlichen Amplitudenfehlers und eines Bildpunktversatzes für ein Bild mit 127×127 Bildpunkten angegeben [ANSI01]. Obwohl diese Fehler die subjektive Wahrnehmung kaum oder gar nicht stören, verursachen sie nur geringe PSNR-Werte.

2.4.2 Merkmalsbasierte Maße

Viele Verfahren zur objektiven Bewertung von Bewegtbild versuchen, Qualitätsmerkmale zu finden, diese zu messen und entsprechend der subjektiven Wahrnehmung zu einem Qualitätsmaß zusammenzuführen. Je nach Anwendung werden verschiedenste Qualitätsmerkmale (z. B. Rauschen, Blockausfälle, Unschärfe) mit oder ohne Referenz gemessen. Es kommen sowohl deterministische als auch statistische Methoden zum Einsatz. Die damit ermittelten Qualitätsfaktoren können mittels einer Regressionsanalyse unter Nutzung eines Faktorenmodells zu einem Qualitätsmaß kombiniert werden. Für die Anpassung des Faktorenmodells ist eine große Anzahl von Testsequenzen subjektiv zu bewerten (siehe Unterabschnitt 2.3.7).

Im Laufe der Entwicklung wurde eine Vielzahl an merkmalsbasierten Verfahren zur Bewertung von Bewegtbild entwickelt [Fec00], [Sar02], [VQEG03] u. a. Neben allgemeinen Verfahren gibt es auch viele Verfahren für spezielle Einsatzzwecke; z. B. Verfahren zur Messung der Unschärfe ohne Referenz [Mar02]. Die Frage, ob es ein universelles Verfahren zur allgemeinen Bewertung der Videoqualität geben kann, wird unterschiedlich beantwortet. Die VQEG (Video Quality Experts Group) hat das Ziel, objektive Verfahren zur Bewertung von Fernsehbildern zu vergleichen und zur Standardisierung zu empfehlen [VQEG03]. Wegen der Komplexität des visuellen Systems ist es jedoch fraglich, ob ein einziges Qualitätsmodell für alle Anwendungen geeignet sein kann [Fec00].

Im Folgenden wird stellvertretend für merkmalsbasierte Videoqualitätsmaße das VQM des ITS vorgestellt, welches auf den Empfehlungen nach ANSI T1.801 basiert.

Begriffe und Parameter für die objektive Bewertung nach ANSI T1.801

In den Standards ANSI T1.801.01 bis T1.801.03 des American National Standards Institute werden Begriffe, Parameter und Testmaterial zur objektiven Bewertung von Video definiert [ANSI95], [ANSI96a], [ANSI96b]. Es wird zwar kein objektives Qualitätsmaß an

sich vorgeschlagen, aber notwendige Grundlagen für dessen Entwicklungen. Da verschiedene Qualitätsmaße diese Grundlagen nutzen, werden sie hier kurz vorgestellt.

In ANSI T1.801.01-1995 werden Testszenen zur objektiven und subjektiven Qualitätsbewertung spezifiziert [ANSI95]. Diese wurden u. a. von ITU-R BT.802-1 [ITU94b] übernommen. In ANSI T1.801.02-1996 werden Begriffe zur Beschreibung von digitalem Video und dessen Qualität definiert [ANSI96a]. Dazu gehören u. a.: *video, block distortion, blurring, mosquito noise, edge business, jerkiness, error blocks*.

In ANSI T1.801.03-1996 werden Merkmale und Parameter als Basis zur Bewertung von digitalem Video spezifiziert [ANSI96b]. Eine Gruppe von Parametern bezieht sich auf synthetische Testsignale (vergleichbar mit den Testzeilen des analogen Fernsehens) zur Bestimmung von Kontrastfehlern, Bildversatz u. ä. Die Merkmale und Parameter für natürliche Vorlagen werden unterteilt in Skalare, Vektoren und Matrizen. Zur Bestimmung des örtlichen Informationsgehaltes (*spatial information, SI*) werden Sobel-Hochpassfilter in vertikaler und horizontaler Richtung angewendet (siehe Gleichungen 2.20 und 2.21):

$$H_v = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}, \quad H_h = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad (2.20)$$

$$SI_v(i, j, t_n) = Y(i, j, t_n) \otimes H_v, \quad SI_h(i, j, t_n) = Y(i, j, t_n) \otimes H_h \quad (2.21)$$

Der Betrag und die Richtung der örtlichen Information ergeben sich zu:

$$SI_r(i, j, t_n) = \sqrt{SI_v(i, j, t_n)^2 + SI_h(i, j, t_n)^2}, \quad SI_\Theta(i, j, t_n) = \arctan \frac{SI_v(i, j, t_n)}{SI_h(i, j, t_n)} \quad (2.22)$$

Relevant für die Qualitätsbewertung sind die statistischen Maße der örtlichen Information: arithmetisches Mittel, Standardabweichung und quadratisches Mittel.

$$SI_{mean}(t_n) = \frac{1}{P} \sum_i \sum_j SI_r(i, j, t_n)$$

$$SI_{stdev}(t_n) = \sqrt{\left[\frac{1}{P} \sum_i \sum_j SI_r^2(i, j, t_n) \right] - SI_{mean}^2(t_n)}$$

$$SI_{rms}(t_n) = \sqrt{SI_{mean}^2(t_n) + SI_{stdev}^2(t_n)} \quad (2.23)$$

Des Weiteren kann die örtliche Information mit $HV(r_{min}, \Delta\Theta, t_n)$ in horizontaler und vertikaler Richtung ausgewertet werden. Mit $\Delta\Theta$ wird die für die Summation maximal zulässige Richtungsabweichung von der Horizontalen oder der Vertikalen beschrieben. Der für die Summation minimal zulässige Betrag $SI_r(i, j, t_n)$ wird mit r_{min} angegeben.

$$HV(r_{min}, \Delta\Theta, t_n) = \frac{1}{P} \sum_i \sum_j SI_r(i, j, t_n)$$

mit $SI_r(i, j, t_n) \geq r_{min} > 0$

$$\text{und } k \cdot \frac{\pi}{2} - \Delta\Theta < SI_\Theta(i, j, t_n) < k \cdot \frac{\pi}{2} + \Delta\Theta \quad (k = 0, 1, 2, 3) \quad (2.24)$$

Zur Bestimmung des zeitlichen Informationsgehaltes (*temporal information, TI*) wird die Differenz der Luminanz eines Bildpunktes zu der im vorherigen Bild bestimmt:

$$TI(i, j, t_n) = Y(i, j, t_n) - Y(i, j, t_{n-1}) \quad (2.25)$$

Wie im Falle von SI werden die statistischen Größen bestimmt:

$$\begin{aligned}
 TI_{mean}(t_n) &= \frac{1}{P} \sum_i \sum_j TI(i, j, t_n) \\
 TI_{stdev}(t_n) &= \sqrt{\left[\frac{1}{P} \sum_i \sum_j TI^2(i, j, t_n) \right] - TI_{mean}^2(t_n)} \\
 TI_{rms}(t_n) &= \sqrt{TI_{mean}^2(t_n) + TI_{stdev}^2(t_n)}
 \end{aligned} \tag{2.26}$$

Das vektorbasierte Merkmal $\mathbf{f}(t_n)$ enthält die Fourier-Transformierten einer quadratischen Subregion, wobei die Koeffizienten F der gleichen Ortsfrequenz aufsummiert werden.

$$\begin{aligned}
 f(k, t_n) &= \frac{1}{N_k} \sum_{i,j} F(i, j, t_n) \quad \text{für alle } k-1 < \sqrt{i^2 + j^2} \leq k \\
 \mathbf{f}(t_n) &= \begin{bmatrix} f(0, t_n) \\ f(1, t_n) \\ \vdots \\ f(\frac{N}{2} - 1, t_n) \end{bmatrix}
 \end{aligned} \tag{2.27}$$

Als matrixbasiertes Merkmal wird die Matrix $\mathbf{Y}'_{out}(t_n)$ der Luminanzwerte aller Bildpunkte eines Bildes spezifiziert. Helligkeit (Gleichspannungsanteil I_Y) und Kontrast (Verstärkungsfaktor g_Y) können zuvor korrigiert worden sein.

$$\mathbf{Y}'_{out}(t_n) = \frac{\mathbf{Y}_{out}(t_n) - I_Y}{g_Y} \tag{2.28}$$

Im Standard sind weiterhin verschiedene Funktionen zum Vergleich der genannten Größen a (z. B. TI , SI_{stdev}) spezifiziert. Dies sind u. a. die Minima, Maxima und quadratischen Mittel von $\log_ratio(t_n)$ und $err_ratio(t_n)$

$$\begin{aligned}
 \log_ratio(t_n) &= \lg \frac{a_{out}(t_n)}{a_{in}(t_n - d)} \\
 err_ratio(t_n) &= \frac{a_{in}(t_n - d) - a_{out}(t_n)}{a_{in}(t_n - d)}
 \end{aligned} \tag{2.29}$$

Weitere Vergleichsfunktionen dienen zur Ermittlung von Änderungen im Ortsfrequenzspektrum, Änderungen des PSNR etc. Durch die Vergleichsfunktionen ist eine Vielzahl an Parametern P möglich, z. B. das *maximum lost edge energy*:

$$P = \max \left\{ err_ratio(t_n) = \frac{SI_{stdev,in}(t_n - d) - SI_{stdev,out}(t_n)}{SI_{stdev,in}(t_n - d)} \right\}_{n=0}^{M-1} \tag{2.30}$$

Letztlich werden im Standard Bezüge zwischen den Parametern und den Störungen genannt, z. B. ist das *maximum lost edge energy* mit Verwischungen und Verschmierungen (*blurring*, *smearing*) assoziiert.

VQM von ITS

Am Institute for Telecommunication Services (ITS) des Wirtschaftsministeriums der USA wurden mehrere Modelle zur objektiven Qualitätsbeurteilung von digitalem Video entwickelt [Wol02]. Die verschiedenen Modelle basieren auf den gleichen Prinzipien, sind aber für unterschiedliche Anwendungen ausgelegt (z. B. allgemeine Nutzung, TV, Video-Konferenzen). Es werden perceptionsorientierte Merkmale ermittelt, die je nach Modell unterschiedlich gewichtet zusammengeführt werden. In Abbildung 2.10 sind die einzel-

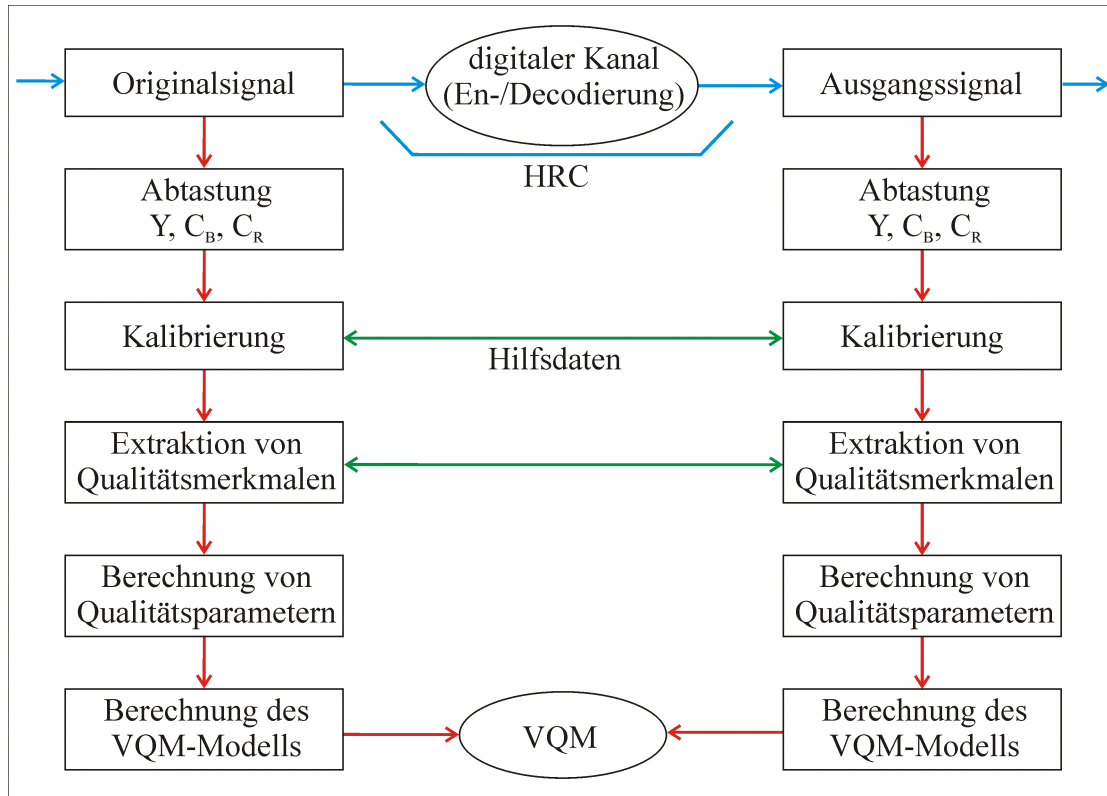


Abbildung 2.10: Schritte zur Bestimmung des VQM; nach [Wol02]

nen Schritte zur Ermittlung des VQM nach ITS skizziert. Das VQM basiert auf einem Vergleich mit dem Originalsignal, sodass dieses für die Berechnungen vollständig zur Verfügung stehen muss. Dieses hat in den drei Komponenten der Luminanz Y und den Farbdifferenzen C_B , C_R gemäß ITU-R BT.601 [ITU94a] vorzuliegen. Die anschließende Kalibrierung ist notwendig, damit die zu bestimmenden Qualitätsparameter bildpunktweise bestimmt werden können. Sie besteht aus den vier Schritten:

1. Ermittlung und Korrektur von örtlichen Bildverschiebungen
2. Bestimmung eines für die Bewertung gültigen Bildausschnittes
3. Abschätzung und Korrektur von Helligkeit und Kontrast
4. Ermittlung und Korrektur von Verzögerungen des Videosignals

Die Korrekturen werden selbstverständlich nur am zu bewertenden Signal durchgeführt, während Schritt 2 auch für das Originalsignal notwendig ist. Die Ermittlung von Bildverschiebungen, von Kontrast- und Helligkeitsfehlern sowie von Verzögerungen erfolgt iterativ.

Für die Suche nach örtlichen Verschiebungen wird zunächst ein grobes Raster (typ. vier oder acht Bildpunkte) und anschließend ein feines Raster (bildpunktweise) verwendet. Als Suchkriterium wird das Minimum der Standardabweichung des Differenzbildes genutzt.

Ein für die Bewertung gültiger Bildausschnitt wird bestimmt, um eine unzulässige Bewertung von schwarzen Rändern oder von durch Bildverschiebungen bedingten Rändern auszuschließen.

Die Korrektur von Helligkeit und Kontrast erfolgt für Y , C_B und C_R separat. Der Kontrast entspricht hier einem Verstärkungsfaktor g und die Helligkeit entspricht einem Gleichspannungsanteil l . Für die Matrizen des zu messenden Bildes \mathbf{P} und des Originalbildes \mathbf{O} gilt dann:

$$\mathbf{P} = g \cdot \mathbf{O} + l \quad (2.31)$$

Dieses Gleichungssystem ist für n Bildpunkte $(n - 1)$ -fach überbestimmt. Es wird mit der Methode der kleinsten Quadrate gelöst.

Für die Ermittlung von Verzögerungen des Videosignals werden die Parameter $TI2$, $TI4$, Y_{mean} und $TI10$ genutzt. Die TI -Parameter entsprechen den Parametern für die temporale Information nach ANSI T1.801.03 für den Abstand von zwei, vier und zehn Halbbildern. Y_{mean} ist das arithmetische Mittel aller Luminanzwerte der Sequenz. Für die Abschätzung der zeitlichen Verzögerung werden zunächst $TI2$ und dessen Standardabweichung benutzt. Liegen sie unterhalb definierter Schwellenwerte, wird $TI4$ genutzt. Falls auch $TI4$ nicht verlässlich ist, wird auf Y_{mean} und gegebenenfalls auf $TI10$ zurückgegriffen.

Im Anschluss an die Kalibrierung werden die Qualitätsmerkmale bestimmt. Diese werden aus den folgenden Größen bestimmt:

- die Standardabweichung der räumlichen Information SI einer S-T-Region (z. B. 8×8 Bildpunkte, 6 Bilder); nach Filterung mit einem 13×13 -Sobel-ähnlichen Filter
 - horizontal/vertikal⁵: f_{SI13}
 - diagonal: f_{HI13}
- die mittlere euklidische Distanz der Chrominanzvektoren C_B , C_R innerhalb einer S-T-Region (z. B. 4×4 Bildpunkte, 3 Bilder): f_{COHER_COLOR}
- die Standardabweichung der Luminanz Y in einer S-T-Region (\rightarrow Kontrast): f_{CONT}
- die Standardabweichung der temporalen Information in einer S-T-Region: f_{ATI}
- das Produkt aus den Standardabweichungen von Luminanz und temporaler Information: $f_{CONTRAST_ATI}$

Für die Bildung der Qualitätsmerkmale werden die genannten Größen für Original- und Messequenz mittels verschiedener Verfahren in Beziehung gesetzt:

- absolutes oder logarithmisches Verhältnis (Gewinn oder Verlust): $ratio_loss(s, t)$, $ratio_gain(s, t)$, $log_loss(s, t)$, $log_gain(s, t)$
- euklidische Distanz: $euclid(s, t)$

⁵vergleiche [ANSI96b]

- arithmetisches Mittel oder Standardabweichung der Größen; bezogen auf Zeit oder Ort: *mean*, *std*
- nonlineare Skalierung: *square*, *sqrt* (Exponent 2 bzw. 0,5)
- Begrenzung des Wertebereiches (Clipping), z. B. 10% oder 50% der Werte über dem definierten Schwellenwert

Das ITS hat zunächst fünf VQM-Modelle definiert (*Television*, *Videoconferencing*, *General*, *Developer* und *PSNR*). Diese errechnen sich als gewichtete Summe einer Auswahl von für das jeweilige Modell relevanten Qualitätsparamter p_q :

$$VQM_x = \sum_{i=1}^n k_i \cdot p_q \quad (2.32)$$

Als Beispiel sei hier das Modell *Developer* genannt:

$$\begin{aligned} VQM_D = & -0,6289 \cdot avg18F_Y_si13_8 \times 8_std_6_ratio_loss_below5\%_mean_clip_0,03 \\ & +0,2305 \cdot avg18F_Y_hv13_angle0,225_rmin20_8 \times 8 \\ & \quad _mean_3_ratio_loss_below5\%-10\%_square_clip_0,06 \\ & +0,1551 \cdot avg18F_Y_hv13_angle0,225_rmin20_8 \times 8 \\ & \quad _mean_3_log_gain_above95\%_mean \\ & +1,0587 \cdot avg18F_Y_ati_8 \times 8_std_1_log_gain_mean_10\% \\ & +0,1444 \cdot avg18F_Y_ati_8 \times 8_std_3_ratio_loss_below_5\%-10\% \end{aligned} \quad (2.33)$$

2.4.3 HVS-basierte Maße

Ein anderer Ansatz zur Qualitätsbewertung von Bewegtbild ist die Nachbildung des menschlichen visuellen Systems (HVS). Hierbei wird versucht, das Bewegtbild in gleicher Weise wie der Visus zu verarbeiten. Die Funktionsweise des Visus ist sehr komplex, da sie sich nicht ausschließlich durch Amplitudenkennlinien und Ortsfrequenzgänge beschreiben lässt [Hau94]. Die Wahrnehmungseigenschaften können durch andere Stimuli maskiert werden. So werden beispielsweise Störungen in Vordergrundobjekten durch bewegte Hintergründe vom Visus maskiert. Mehrkanalige HVS-Modelle versuchen solche Interaktionswirkungen in Abhängigkeit von der Frequenz und der Orientierung der Stimuli zu berücksichtigen [Fec99], [Hau94]. Im Folgenden werden zwei HVS-Modelle vorgestellt.

Heuristisches optimiertes HVS-Modell nach Fechter

Ein Vorschlag zur Nachbildung des HVS wurde von Fechter im Jahre 1998 vorgestellt [Fec98], [Fec99]. Dieses Einkanal-Qualitätsmodell für komprimierte Bildfolgen wurde heuristisch optimiert. Insbesondere wurde die Beurteilung von Störartefakten oberhalb der Wahrnehmungsschwelle optimiert, damit die Bewertung über den gesamten Messbereich der subjektiven Wahrnehmung entspricht. In Abbildung 2.11 ist ein Blockbild des entwickelten Qualitätsmodells dargestellt. Die Korrektur der Nichtlinearitäten dient zum Ausgleich der Charakteristik der Bildwiedergaberöhre und der menschlichen Helligkeitswahrnehmung, die eine logarithmische Charakteristik hat (Weber-Fechner-Gesetz).

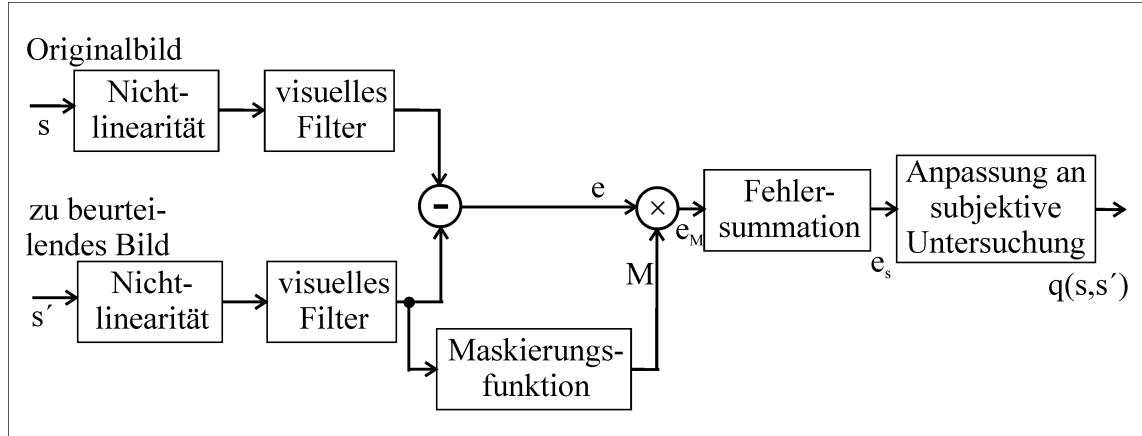


Abbildung 2.11: Qualitätsmodell nach Fechter; [Fec98]

Das visuelle Filter repräsentiert den Ortsfrequenzgang des HVS. Dieser Frequenzgang wurde mit $A(f)$ angenähert (f in c/deg - cycles per degree).

$$A(f) = (1 + b \cdot f) \cdot e^{-c \cdot f} \quad \text{mit } b, c > 0 \quad (2.34)$$

Für die Anpassung des Modells an das HVS wird das Fehlersignal e (Differenz zwischen Original- und Messsignal) maskiert. Die Wahrnehmung von Störungen im Bereich von Signalsprüngen ist wesentlich kleiner als in homogenen Flächen. Die Detektion von Signalsprüngen erfolgt mit einer Aktivitätsfunktion A des betrachteten Bildpunktes $s(x, y)$:

$$A = \max(|s(x, y) - s(x, y + 1)|, |s(x, y) - s(x, y - 1)|, |s(x, y) - s(x + 1, y)|, |s(x, y) - s(x - 1, y)|) \quad (2.35)$$

Die Maskierungsfunktion ist:

$$M(A) = \frac{1}{1 + v \cdot A} \quad (2.36)$$

Durch die Multiplikation der Maskierungsfunktion M mit dem Fehlersignal e werden Bildanteile mit hoher Aktivität gedämpft, d. h. für die Fehlersummentation weniger berücksichtigt. Bei der anschließenden Fehlersummentation werden die einzelnen Fehler des Bildes zusammengefasst. Die Fehlersummentation wird hierbei als Zwischenlösung zwischen einer Maximal- und einer Mittelwertbildung ausgeführt. Es wird jeweils innerhalb eines quadratischen Fensters 4×4 der mittlere Fehler bestimmt. Das Maximum dieser mittleren Fehler dient als Qualitätsmaß.

$$S = \max \sqrt{\frac{1}{n^2} \sum_{x=i+1}^{i+n} \sum_{y=j+1}^{j+n} (s(x, y) - s'(x, y))^2} \quad (2.37)$$

Im letzten Schritt erfolgt die Berücksichtigung einer temporalen Eigenschaft der subjektiven Bewertung. Dies ist der so genannte *memory effect*, welcher den Probanden in seiner Bewertung von den vorhergehenden Sequenzen beeinflussen lässt. Dieser Einfluss wird mit einem Exponentialfilter $g(n)$ nach ITU-R BT.500 [ITU02] korrigiert.

$$g(n) = \sum_{j=n}^{n-m} e_s(j) \cdot e^{-\frac{j}{T}} \quad (2.38)$$

Zur Anpassung des Ausgabewertes an die Ausgabewerte der subjektiven Messverfahren wird dieser mittels linearer Regression auf eine Skala von 0...100 skaliert.

$$\hat{q}(s, s') = a + b \cdot g \quad (2.39)$$

Das gesamte Einkanal-Qualitätsmodell wurde von Fechter mit Hilfe subjektiver Tests heuristisch optimiert [Fec98]. Da es sich um die Beurteilung von Störungen digitaler Videosequenzen handelt, wurden SSCQE-Tests nach ITU-R BT.500 [ITU02] durchgeführt. Mit den gewonnenen Testergebnissen wurden zunächst die Zeitkonstante T und der Offset τ des Glättungsfilters bestimmt. Diese beeinflussen die Korrelation mit dem MOS in einem Wertebereich von ca. 0,1. Die Korrektur der Nichtlinearitäten im ersten Schritt des Einkanal-Modells ist nicht erforderlich, da sie einen Korrelationsgewinn von nur 0,005 bewirkt. Die Optimierung des Koeffizienten c des visuellen Filters beeinflusst den Korrelationskoeffizienten um ca. 0,08. Die Korrelation ist für $c = 0,3$ und $b = 0$ maximal, d. h. wenn der visuelle Filter als Tiefpassfilter arbeitet. Der Koeffizient v der Maskierungsfunktion hat einen Einfluss auf die Korrelation in der Größenordnung 10^{-2} und ist mit 0,06 optimal. Schließlich wurde die Fenstergröße der Summationsfunktion als Parameter optimiert. Sie ist für $n = 4$ optimal.

Das „JNDmetrix“-Modell

Das „JNDmetrixTM“-Maß wurde vom Unternehmen Sarnoff Corporation (USA) entwickelt [Sar02]. Es beruht auf einem proprietären HVS, dem „JNDmetrix Human Perception Model“ [Sar01]. Das verwendete Maß für die JND (*Just-Noticeable Difference*) ist wie folgt definiert:

- 1 JND entspricht einem Unterschied, der von 75% der Probanden erkannt wird.
- Die höheren JND-Werte sind durch das Inkrement definiert. Wenn $JND(X,Y)=1$ und $JND(Y,Z)=1$, dann ist $JND(X,Z)=2$.

Dies bedeutet für einen JND-Wert von 2, dass $(0,75 + 0,75 \cdot (1 - 0,75)) = 93,75\%$ der Probanden einen Unterschied feststellen. [Sar01]

JND-Wert	1	2	3	4	5
Wahrscheinlichkeit in %	75	93,8	98,4	99,6	99,9

Tabelle 2.4: Wahrscheinlichkeiten der Unterscheidung für das JND-Maß [Jan02]

< 1 JND	Die Bildsequenzen können (selbst mit Wissen) nicht unterschieden werden.
1 JND	Unterschiede sind auch für den geübten Betrachter kaum apparent.
3 JND	Die Unterschiede sind bei genauer Betrachtung apparent.
5 JND	Die Unterschiede sind deutlich apparent.

Tabelle 2.5: Erläuterungen zu JND-Werten [Sar01]

Zur Qualitätsbewertung wird zunächst für jeden Bildpunkt des zu bewertenden Bewegtbildes ein JND-Wert bestimmt. Diese JND-Karten (*maps*) werden mit Hilfe des proprietären HVS für die Luminanz und für die Chrominanz bestimmt. In Abbildung 2.12 ist die Abfolge der JND-Berechnung dargestellt [Rav98]. In einem ersten Schritt erfolgt eine Wandlung der RGB-Werte und die Anpassung der YUV-Werte an die psychooptische

Gradation. Im nächsten Schritt erfolgt eine Zerlegung in Gauß-Pyramiden, sodass die Bildsequenzen für die Kontrastbestimmung für unterschiedliche Auflösungen erfolgen kann. Die Normierung der Luminanzwerte dient zur Nachbildung der Adaption des Visus. Auf Basis der normierten Werte werden das Richtungskонтраст- und das Flimmerkontrastempfinden nachgebildet. Der Farbkонтраст wird mittels der Chrominanzwerte bestimmt. Im letzten Schritt wird die Maskierung des Kontrastempfindens in bewegten Bildbereichen u. a. simuliert. Die Ergebnisse dieser Prozesse liegen in Form von JND-Maps für die Luminanz und die Chrominanz vor.

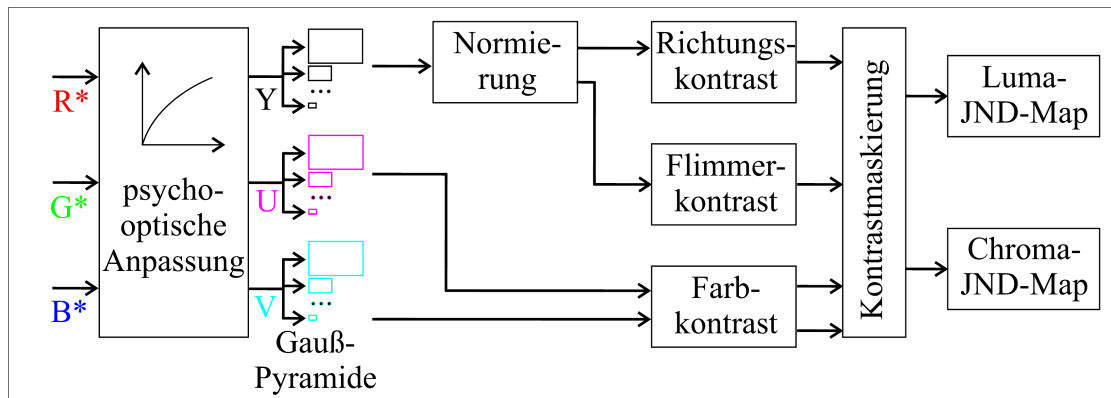


Abbildung 2.12: Bestimmung der JND-Maps

Die JND-Maps der Original- und der Testsequenz werden mit einem Differenzoperator verknüpft und diese Differenzen werden über alle Pixel, Pyramidenebenen, Kanäle und Bilder aufsummiert [Rav98]. Das Ergebnis bildet das „JNDmetrixTM“-Maß.

2.4.4 Bewertung der Segmentierungsqualität

Der objekt- und szenenbasierte Ansatz zur Codierung audiovisueller Inhalte etablierte sich mit der Standardisierung von MPEG-4 [ISO98], [Sik03]. Als Quelle für beliebig geformte Videoobjekte (*arbitrarily shaped video objects*) dienen meist konventionelle Videoaufnahmen, die hierfür segmentiert werden müssen. Ein verbreitetes Beispiel für eine Segmentierungsmethode sind Fernsehstudioaufnahmen vor einer blauen Wand, die mittels Chroma-Keying-Technik gestanzt werden. Mit der Einführung von Videoobjekten wuchs auch der Bedarf an geeigneten Segmentierungsverfahren für beliebige Aufnahmen. Damit einher ging die Frage nach der Bewertung der Segmentierungsqualität für Multimedia-Anwendungen [Erd01], [Cor01], [Cor03], [Vil99]. Untersuchungen zur Segmentierungsqualität wurden zuvor bereits im Rahmen der Bildverarbeitung (Objekterkennung) vorgenommen. Da die Objekterkennung andere Anforderungen stellt, sind diese Qualitätsmaße nicht oder nur ansatzweise anwendbar.

Die Bewertung der Segmentierungsqualität kann nach unterschiedlichen Gesichtspunkten erfolgen [Cor03]:

- *individuelle Objektsegmentierungsqualität*: Die Qualität der Segmentierung eines einzelnen Objektes wird unabhängig von der Gesamtszene bewertet.
- *gesamte Segmentierungsqualität*: Die Qualität der Segmentierung aller Objekte einer Szene wird in ihrer Gesamtheit bewertet.

Ebenso steht die Frage nach einer Referenz, d. h. ob mit einer Referenzsegmentierung verglichen werden kann [Cor03]:

- *alleinstehende Bewertung*: Eine Referenzsegmentierung ist nicht verfügbar.
- *relative Bewertung*: Die Segmentierung kann mit einer Referenzsegmentierung verglichen werden.

Die Verfügbarkeit einer Referenzsegmentierung hat entscheidenden Einfluss auf die Auswahl der Bewertungskriterien und -methoden.

Relative Bewertung

In [Erd01], [Cor03], [Vil99] werden verschiedene Maße zur relativen Bewertung der Segmentierungsqualität vorgestellt.

Für die Bewertung der örtlichen Genauigkeit werden nach [Vil99] die Anzahl der fehlenden Vordergrundbildpunkte (*missing foreground points*, MF) und die Anzahl der dem Hintergrund falsch zugeordneten Bildpunkte (*added background points*, AB) zonenweise gemessen und zu einem örtlichen Qualitätsmaß (*spatial quality measure*, SQM) aufsummiert. Die Wichtungsfunktionen w_{MF} und w_{AB} dienen dazu, fehlerhaft zugeordnete Zonen entsprechend ihrem Abstand d zur Objektgrenze stärker zu wichten:

$$SQM = \sum_{d=1}^{D_{FGmax}} w_{MF}(d) \cdot \#(R_d \cap E^C) + \sum_{d=1}^{D_{BGmax}} w_{AB}(d) \cdot \#(R_d^C \cap E) \quad (2.40)$$

Die zeitliche Kohärenz (*variation of SQM*, vQM) wird zunächst über die zeitliche Variation des SQM bestimmt:

$$vQM(t) = |Q_{MF}(t) - Q_{MF}(t-1)| + |Q_{AB}(t) - Q_{AB}(t-1)| \quad (2.41)$$

Des Weiteren wird Variation des Gravitationszentrums VGC der Objektform bestimmt:

$$VGC(t) = \left\| \left(\vec{GC}^{est}(t-1) - \vec{GC}^{ori}(t-1) \right) - \left(\vec{GC}^{est}(t) - \vec{GC}^{ori}(t) \right) \right\| \quad (2.42)$$

Ähnliche Ansätze werden in [Erd01] genutzt. Zur Bestimmung der Segmentierungsqualität werden hier vier Distanzmaße bestimmt. Fehlzuordnungen werden mittels der *misclassification penalty* erfasst:

$$D_i^t = \frac{\sum_{(k,l) \in (s_i \cup g_i)} I^t(k,l) \cdot w_{gi}(k,l)}{\sum_{all(k,l)} w_{gi}(k,l)} \quad (2.43)$$

Hierbei ist I^t die Indikatorfunktion, ob der Bildpunkt (k,l) in der Referenz die gleiche Zuordnung aufweist. Die Bildpunkte der zu bewertenden Segmentierung werden mit s_i und der Referenz (*ground truth*) mit g_i angegeben. Die Funktion w_{gi} ist ein modifiziertes Chamfer-Abstandsmaß. Ein weiteres Maß ist die *shape penalty*, welche aus den Differenzen der Tangentialwinkel $\Theta_{gi}^t(s)$ und $\Theta_{si}^t(s)$ von Referenz und zu bewertender Segmentierung gebildet wird:

$$DS^t = \frac{1}{M} \sum_{i=1}^M \frac{\sum_{s=1}^P |\Theta_{gi}^t(s) - \Theta_{si}^t(s)|}{P \cdot 2 \cdot \pi} \quad (2.44)$$

Zur Messung der zeitlichen Kohärenz werden *misclassification penalty* und *shape penalty* zeitlich gemittelt. Ein weiteres Maß ist die *motion penalty*. Hierfür werden die Bewegungsvektoren M_{gi}^t und M_{si}^t der Referenz und zu bewertenden Segmentierung verglichen:

$$D_i = \frac{1}{T} \sum_{t=1}^T \frac{\|M_{gi}^t - M_{si}^t\|}{\|M_{gi}^t\| + \|M_{si}^t\|} \quad (2.45)$$

Zur Gesamtbeurteilung wird die *combined penalty* als Kombination der Einzelmaße gebildet:

$$CP = \alpha \cdot DP + \beta \cdot DS + \gamma \cdot DM \quad (2.46)$$

Die vorgestellten Maße zeigen die Ähnlichkeiten in ihren Ansätzen. In [Cor03] werden die Maße der verschiedenen Verfahren zur relativen Segmentierungsbewertung klassifiziert.

- *örtliche Genauigkeit*: Genauigkeit der Form, Übereinstimmung der örtlichen Merkmale, Anzahl der segmentierten Objekte
- *zeitliche Kohärenz*: zeitliche Stabilität

Ebenso wird die Eignung von Videoqualitätsmaßen (SI , TI) zur Bewertung der Segmentierungsqualität festgestellt.

Alleinstehende Bewertung

In der Mehrzahl der Anwendungen muss die Segmentierungsqualität ohne Referenz ermittelt werden, was die Bildung geeigneter Maße erheblich erschwert. Man kann sich behelfen, indem man Objekteigenschaften festlegt, von denen man eine Beeinflussung der Qualität erwartet. Hierfür wurde in [Cor03] die Homogenität eines Objektes (*intra-object homogeneity*) mittels Eigenschaften, wie Gleichmäßigkeit der Objektform und Gleichförmigkeit sowie zeitlicher Stabilität und Gleichförmigkeit der Bewegung beschrieben. Die Gleichmäßigkeit der Objektform wird mit Hilfe geometrischer Größen erfasst:

$$Kompaktheit = \frac{u^2(E)}{A(E)} \quad Rundheit = \frac{4 \cdot \pi \cdot A(E)}{u^2(E)} \quad Ausdehnung = \frac{A(E)}{(2 \cdot D(E))^2}$$

mit: u : Umfang, A : Flächeninhalt, D : Dicke (2.47)

Zur Messung der Gleichförmigkeit des Objektes dienen der örtliche Informationsgehalt SI und die Varianz. Die zeitliche Stabilität wird durch die zeitliche Änderung der Merkmale: Größe, Ausdehnung und Komplexität (SI/TI) beschrieben. Die Gleichförmigkeit der Bewegung wird als Varianz der Bewegungsvektoren des Objektes berechnet.

Für eine Bewertung ohne Referenz sind grundsätzlich Einschränkungen zu treffen. Die in [Cor03] gesetzten Einschränkungen sind für natürliche Objekte (z. B. Menschen) in typischen Bewegungsabläufen und in typischen Aufnahmen zutreffend und nachgewiesen.

2.4.5 Systematisierung der Bewertungsmaße

Klassifizierung

Die Verfahren zur objektiven Bewertung von digitalen Bildsequenzen basieren auf unterschiedlichen Ansätzen und sind für unterschiedliche Anwendungen konzipiert, sodass eine Klassifizierung in Bezug auf nur ein Kriterium misslingt. Typische Anwendungen sind:

- *Beispiel 1*: Es soll die Codiereffizienz eines Videocodierungsverfahrens ermittelt werden, d. h. welche Videoqualität liefert der Encoder bei einem bestimmten Kompressionsverhältnis. Für die Tests stehen Original- und Messsequenzen vollständig zur Verfügung und es besteht keine Realzeitanforderung.
- *Beispiel 2*: Es soll die Übertragung eines Videodatenstromes überwacht werden. Es können nur einige wichtige Parameter der Originalbildsequenz zum Messpunkt übermittelt werden. Es besteht die Realzeitanforderung.
- *Beispiel 3*: Es sollen bestimmte Fehlertypen (z. B. Blockausfälle) überwacht werden. Hierfür ist das Original nicht zwangsläufig erforderlich und oft besteht die Realzeitanforderung.

Hieraus wird deutlich, dass die vollständige bzw. teilweise Verfügbarkeit einer Referenz ein Klassifizierungsmerkmal sein kann. Nach [ANSI01], [ITU01], [Sar02] u. a. unterscheidet man drei Methoden:

- *Full Reference (FR)*: Diese Methode ist anwendbar, wenn das vollständige Referenzvideosignal zur Verfügung steht. Dies ist eine so genannte *double-ended*-Methode.
- *Reduced Reference (RR)*: Diese Methode ist anwendbar, wenn lediglich einige Informationen über das Referenzvideosignal (z. B. statistische Größen) zur Verfügung stehen. Auch dies ist eine *double-ended*-Methode.
- *No Reference (NR)*: Wenn weder das Referenz-Videosignal noch Informationen darüber zur Verfügung stehen, kann ausschließlich diese Methode angewendet werden. Dies ist eine so genannte *single-ended*-Methode.

Durch diese Klassifizierung werden sehr unterschiedliche Anwendungen und Methoden der Qualitätsbewertung unterschieden. Wenn keine Referenz verfügbar ist, muss eine gänzlich andere Methodik zur Qualitätsbewertung zum Einsatz kommen. Daher ist diese Klassifizierung grundlegend.

Die oben genannten Anwendungen zeigen aber auch, dass unterschiedliche Aspekte für die Bewertung eine Rolle spielen können (z. B. Überwachung von Blockausfällen). Diese Unterschiedlichkeit der relevanten Aspekte wird auch auf Grund der vielen Verfahren und Ansätze deutlich, die für die technische Bildqualitätsmessung bereits entwickelt wurden [ANSI01], [Fec98], [Kay96] u. a. Ein Teil der Verfahren versucht, spezifische Merkmale auszuwerten, während ein anderer Teil mit einem Modell des Visus arbeitet. Dementsprechend wird in [Sar02] folgende Klassifikation vorgeschlagen:

- *Regression-based*: Als Basis für diese Qualitätsmaße dienen statistische Größen, welche aus Referenz- und Testvideo gewonnen werden. Meist werden die statistischen Maße auf Basis subjektiver Testergebnisse zu einem Videoqualitätsmaß verknüpft.
- *Feature-extraction based*: Zur Bewertung werden (erwartete) Störungen extrahiert. Dies betrifft Störungen, die eine abschätzbare subjektive Verschlechterung des Bildes bewirken, z. B. Unschärfe, Rauschen, MPEG-Blockartefakte.
- *Basic-vision calibrated*: Zur Bewertung werden die auftretenden Fehler entsprechend einem visuellen Modell bewertet. Mit solchen Modellen wird das menschliche visuelle

System nachgebildet, indem die Wahrnehmung von Störungen modelliert wird. Diese Wahrnehmung und deren Schwelle sind abhängig vom Ortsfrequenzspektrum der Störung und von deren Umgebung (z. B. bewegter Hintergrund). Dies kann durch Filter und Maskierungsfunktionen im visuellen Modell beschrieben werden.

Die in Abschnitt 2.4.2 erläuterten Maße gehören zur Klasse *regression-based*. Die Methoden der Klasse *Feature-extraction based* ähneln diesen Methoden im Ansatz. Daher wird in anderen Kategorisierungen diese Unterscheidung nicht getroffen. In [Fec99] erfolgt die Unterteilung in Methoden, die das HVS nachbilden und Methoden, die auf Signalextraktion beruhen. In dieser Klassifizierung (siehe Abbildung 2.13) wird die Verwendung einer Referenz für Extraktionsverfahren unterschieden. Die Unterteilung von HVS-Modellen erfolgt in Einkanal- und Mehrkanalmodelle. Die Mehrkanalmodelle versuchen die Wechselwirkungen der visuellen Stimuli (z. B. Maskierung eines Musters durch ein Muster mit ähnlicher Ortsfrequenz) zu erfassen [Fec00].

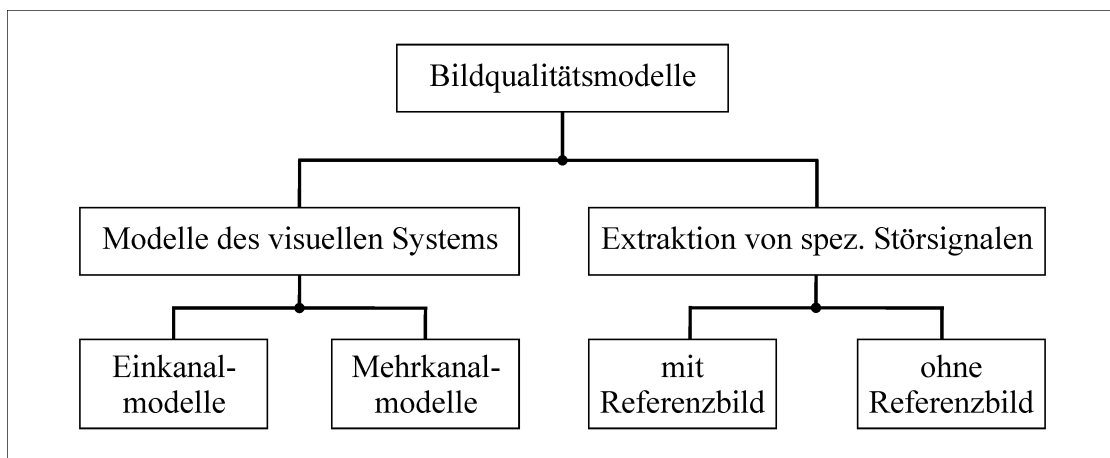


Abbildung 2.13: Methoden zur technischen Bildqualitätsmessung nach [Fec99]

Prinzipien

Die etablierten Verfahren zur Videoqualitätsbewertung orientieren sich naturgemäß an den Prinzipien der Videosignalverarbeitung und -nutzung:

- *Eindeutigkeit des Zustandes*: Konventionelles Bewegtbild weist das Merkmal auf, zu einem bestimmten Zeitpunkt T genau einen festgelegten Zustand zu haben (\cong Einzelbild). Dieses selbstverständliche Merkmal weisen 3D-Videoobjekte nicht auf (siehe Abschnitt 3.1).
- *Uniformität der Repräsentation*: Konventionelles Bewegtbild weist in seiner gesamten Verarbeitungskette (siehe Abbildung 2.1) gleichartige Repräsentationen auf. Beispielsweise kann ein digitales Fernsehbild (DVB) nach der Decodierung am Empfänger mit dem Programmsignal der Fernsehanstalt verglichen werden, da beide Signale gemäß ITU-R BT.601 aufgebaut sind.
- *Wiedergabe*: Video wird typischerweise als Fernsehbild auf Fernsehgeräten oder innerhalb von Multimedia-Anwendungen auf Monitoren wiedergegeben. Das Nutzerverhalten ist eingeschränkt. Dies erleichtert es, der subjektiven Bewertung mittels objektiver Qualitätsbewertungsverfahren nahe zu kommen.

3 3D-Videoobjekte und deren Generierung

3.1 Begriffseingrenzung

Die Vielfalt und Unterschiedlichkeit von Anwendungen, welche dem Begriff *3D-Video* zugeordnet werden, ist immens. Dies begründet sich u. a. aus der unterschiedlichen Nutzung des Begriffes 3D-Video bzw. 3D-Television. Allen Anwendungen ist lediglich gemein, dass die visuelle Wahrnehmung durch räumliche Eindrücke erweitert werden soll. Die folgenden Beispiele verdeutlichen die Vielfalt:

- *Stereoskopie*: Dem Betrachter wird für jedes Auge ein zugehöriges Bild dargestellt, um so einen Eindruck von räumlicher Tiefe zu erzeugen. Der Nutzer schaut mittels einer stereoskopischen Sehhilfe (z. B. einer Polarisationsbrille) auf die Leinwand oder den Monitor.
- *Autostereoskopie*: Als Weiterentwicklung zur Stereoskopie benötigt der Nutzer keine stereoskopische Sehhilfe mehr.
- *visuelle 3D-Wiedergabe*: Darstellung aller drei Dimensionen (z. B. Anzeigeräte basierend auf *Integral Photography*)
- *Blickpunktadaption*: Es wird versucht, den Blickpunkt und die Blickrichtung des Nutzers durch Sensoren zu erfassen und ihm die zugehörige Perspektive zu geben. Man setzt dies zum Beispiel in immersiven Videokonferenzen oder in der Augmented Reality ein.
- *Erfassung des visuellen Raumes*: Das Ziel ist die Repräsentation aller visuellen Eigenschaften einer (natürlichen) Umgebung zur Wiedergabe auf beliebigen Anzeigen (z. B. auf autostereoskopischen Displays, auf konventionellen Monitoren). Ein Verfahren ist z. B. *Ray-Space* (siehe Unterabschnitt 3.2.3).
- *freie Blickrichtungswahl*: betrachterstandpunktfixierte Darstellung mit freier Blickrichtungswahl (omnidirektionales Video)
- *Betrachtung eines 3D-Objektes*: Ein zeitlich veränderliches Objekt kann mit freier Blickpunktwahl betrachtet werden (3D-Videoobjekt).

Diese Beispiele verdeutlichen, dass sich eine Gruppe von Technologien ausschließlich auf die Anzeige bezieht, d. h. die konventionelle Wiedergabe von Bewegtbild wird um einen Tiefeneindruck ergänzt. Eine andere Gruppe von Technologien umfasst neue Formen der Repräsentation und schließt andere Möglichkeiten der Nutzung ein, z. B. der Blickpunkt wird vom Nutzer frei gewählt → Interaktion. Diese Gruppe der 3D-Video-Technologien wurde in der MPEG wie folgt eingeteilt [ISO03a]:

1. omnidirektionales Video (Panoramaansichten)
2. interaktives stereoskopisches Video

3. Free Viewpoint Video (*interactive multiple view video*)4. dreidimensionale Videoobjekte (*free viewpoint video objects*)

Die letztgenannte Kategorie sind die 3D-Videoobjekte (3DVO), die im Folgenden untersucht werden. In [ISO03a] werden sie als Ergebnis einer multiskopischen Aufnahme beschrieben, die durch Aussehen und Form repräsentiert werden.

Ein Unterscheidungskriterium für 3D-Videoobjekte ist, ob sie natürlichen oder synthetischen Ursprungs sind. Natürliche Medienobjekte liegen vor, wenn sie mit einer Kamera oder einem Mikrophon aufgenommen worden sind [ISO98]. Synthetische Objekte sind das Ergebnis einer Modellierung (mittels Rechner). Synthetische 3D-Videoobjekte in diesem Sinne entsprechen demzufolge dem Ergebnis einer Computeranimation eines 3D-Objektes. Dies ist eines der Grundprinzipien der Computeranimation und erfordert daher nicht unbedingt einen eigenständigen Medienobjekttyp. Aus diesem Grund wird meist von natürlichen 3D-Videoobjekten ausgegangen, wenn der Begriff 3D-Videoobjekt verwendet wird. Ebenso sind hier im Folgenden grundsätzlich natürliche 3D-Videoobjekte gemeint. Eine abschließende Eingrenzung des Begriffes 3D-Videoobjekt wird über die Eigenschaften eines 3D-Videoobjektes vorgenommen:

- 3D-Videoobjekte repräsentieren die visuellen Eigenschaften eines natürlichen (dreidimensionalen), zeitveränderlichen Objektes (z. B. eines Menschen).
- Das 3D-Videoobjekt ermöglicht dem Nutzer eine Ansicht auf das Objekt, die von einem wählbaren Standpunkt zu einem wählbaren Zeitpunkt erfolgen kann. Im Idealfall ist der Betrachterstandpunkt einschl. Blickrichtung frei wählbar.



Abbildung 3.1: Zwei Ansichten auf ein SNHC-Szenario mit einem 3D-Videoobjekt

Die Nutzung von 3D-Videoobjekten erfolgt in audiovisuellen Anwendungen, welche auf dem Objekt- und Szenenkonzept basieren. Im einfachsten Fall ist das Objekt singulär. Das Objekt kann jedoch mit anderen Objekten in einer dreidimensionalen Umgebung örtlich und zeitlich verknüpft sein. Die Objekte können gemischt synthetisch und natürlich sein, was zum Prinzip des Synthetic-Natural Hybrid Coding (SNHC) führt [ISO98]. In Abbildung 3.1 sind zwei Ansichten eines solchen SNHC-Szenarios (siehe [Küh03]) dargestellt. Die Person ist durch ein (natürliches) 3D-Videoobjekt repräsentiert. Innerhalb eines limitierten Bereiches erhält der Nutzer während der Navigation eine korrekte Perspektive des Objektes.

3.2 Bildbasierte Verfahren

Für die Generierung von 3D-Videoobjekten sind Ansichten zu generieren, welche nicht mit einer Kamera aufgenommen worden sind. Bei den bildbasierten Verfahren werden hierfür keine dreidimensionalen Modelle der Aufnahmeobjekte rekonstruiert, sondern es werden Interpolations-, Warping- und Morphing-Methoden für die Generierung der Zwischenansichten genutzt. Diese Ansichtengenerierung (*view synthesis*) erfordert zusätzliche Informationen, welche das aufgenommene Objekt beschreiben, z. B. Bildpunktkorrespondenzen, Disparitätenkarten, Tiefenkarten, Visual Hull. In Abbildung 3.2 ist das Ergebnis eines Warping-Prozesses zwischen zwei Aufnahmen dargestellt. Der Vergleich zu den Ausgangsbildern verdeutlicht, dass die View-Synthesis nicht fehlerfrei sein kann.



Abbildung 3.2: Die Interpolation (Mitte) zwischen zwei aufgenommenen Ansichten (links und rechts)

3.2.1 Die Geometrie einer multiskopischen Ansicht

Die bildbasierten Verfahren zur Generierung neuer Ansichten basieren auf den Eigenschaften der multiskopischen Ansicht. Eine solche Ansicht wird mit mindestens zwei (meist stationären) Kameras gewonnen, welche dasselbe Objekt abbilden. Die Abbildung eines Gegenstandspunktes \mathbf{M} ist von dessen Ort und den Parametern der Kameras abhängig. In Abbildung 3.3 ist der Zusammenhang zwischen dem Gegenstandspunkt \mathbf{M} und dessen Abbildungen \mathbf{m} und \mathbf{m}' dargestellt. Die Abbildung \mathbf{m}' liegt dabei immer auf der Abbildung des Strahles von \mathbf{M} nach \mathbf{m} auf die Bildebene E' . Dieser abgebildete Lichtstrahl $\mathbf{l}'_{\mathbf{m}}$ wird *Epipolarlinie* genannt und die Schnittpunkte \mathbf{e} und \mathbf{e}' zwischen der Verbindungsgerade der Fokusse mit den Bildebenen heißen *Epipole*. Die Epipole \mathbf{e} und \mathbf{e}' und Epipolarlinien $\mathbf{l}_{\mathbf{m}}$ und $\mathbf{l}'_{\mathbf{m}}$ liegen in einer Ebene. Die Epipolarlinie $\mathbf{l}'_{\mathbf{m}}$ kann mittels $\mathbf{l}'_{\mathbf{m}} = \mathbf{F}\mathbf{m}$ beschrieben werden, wobei \mathbf{F} eine 3×3 -Matrix ist, welche die Konfiguration der beiden Kameras beschreibt [Fau01], [Sei97], [Müh02]. Diese Matrix \mathbf{F} heißt *Fundamentalmatrix* und erlaubt die Formulierung der Epipolarbedingung:

$$\mathbf{m}'^T \mathbf{F} \mathbf{m} = 0 \quad (3.1)$$

Die Epipolarbedingung beschreibt den grundlegenden Zusammenhang zweier Abbildungen. Da sie unabhängig von \mathbf{M} ist, kann für ein gegebenen Bildpunkt \mathbf{m} bei bekannter Fundamentalmatrix \mathbf{F} die Abbildung \mathbf{m}' bestimmt werden.

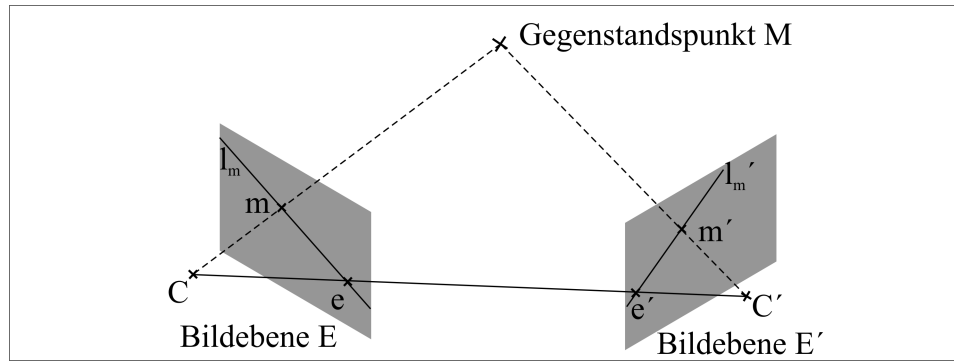


Abbildung 3.3: Die Epipolar-Geometrie zweier Abbildungen

Die Fundamentalmatrix beschreibt die inneren und äußeren Parameter der Kameras und wird für die Synthese neuer Ansichten benötigt. Eine wesentliche Vereinfachung ergibt sich, wenn die Bildebenen parallel zueinander ausgerichtet sind. Die Epipolarlinien sind dann horizontal ausgerichtet und die Fundamentalmatrix wird für die Morphing-basierten View-Synthesis-Verfahren nicht mehr benötigt. Für die Aufnahme von Objekten sind jedoch meist konvergente Kamera-Setups erforderlich. Die Fundamentalmatrix kann mittels einer festen Kalibrierung oder einer Selbstkalibrierung ermittelt werden. Bei einer festen Kalibrierung werden die Kameraparameter bestimmt und daraus die Fundamentalmatrix errechnet. Bei einer Selbstkalibrierung werden in den Abbildungen der Kameras Bildpunkte ermittelt, die denselben Objektpunkt darstellen. Aus Gleichung 3.1 und der Bedingung $\det \mathbf{F} = 0$ lassen sich mit sieben Korrespondenzpunktpaaren die neun unbekannten Elemente der Fundamentalmatrix bestimmen. Jedoch ist das Gleichungssystem vom Grad drei, sodass sich drei Lösungen ergeben. Für die praktische Durchführung ist relevant, dass die gefundenen Korrespondenzpunktpaare nicht grundsätzlich exakt ermittelt werden können. Daher wird die Einhaltung der Epipolarbedingung zu einem Minimierungsproblem, für die das Kriterium entsprechend den geometrischen Auswirkungen gewählt werden muss [Fau01].

Morphing-basierte View-Synthesis-Verfahren sind meist für den eindimensionalen Fall implementiert und erfordern daher horizontale Epipolarlinien, die den Bildzeilen entsprechen [Sei96], [Müh02]. Dies ist für parallele Ansichten der Fall, jedoch können konvergente Ansichten so verzerrt werden, als seien sie parallel aufgenommen worden. Diese Verzerrung heißt *Rektifizierung* (siehe Abbildung 3.4). Die Rektifizierung entspricht einer Reprojek-

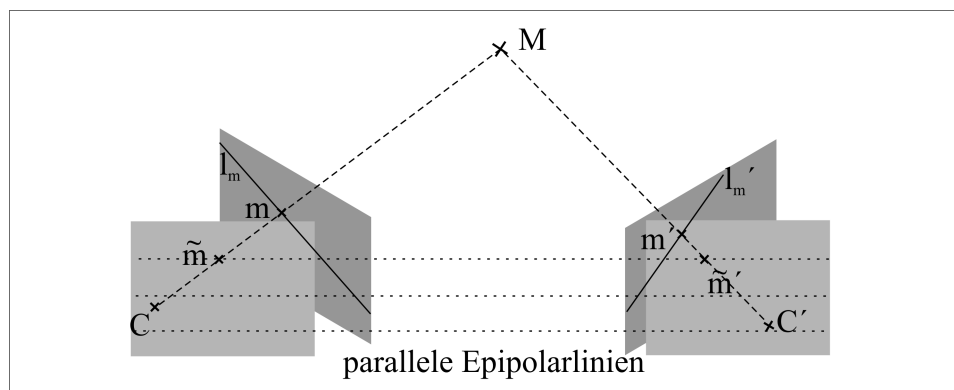


Abbildung 3.4: Rektifizierung

tion, welche algebraisch mit Hilfe der Fundamentalmatrix durchgeführt werden kann. Die Bildpunkte $\tilde{\mathbf{m}}$ und $\tilde{\mathbf{m}}'$ in den rektifizierten Bilder ergeben sich durch Projektion mit den Projektionsmatrizen \mathbf{H} und \mathbf{H}' .

$$\tilde{\mathbf{m}} = \mathbf{H}\mathbf{m} \quad \tilde{\mathbf{m}}' = \mathbf{H}\mathbf{m}' \quad (3.2)$$

Diese Projektionen müssen ebenso die Epipolarbedingung erfüllen:

$$\begin{aligned} \tilde{\mathbf{m}}'^T \cdot \tilde{\mathbf{F}} \cdot \tilde{\mathbf{m}} &= 0 \\ \text{mit: } \tilde{\mathbf{m}}'^T &= \mathbf{m}'^T \mathbf{H}'^T \\ \tilde{\mathbf{F}} &= \mathbf{H}'^{-T} \mathbf{F} \mathbf{H}^{-1} \\ \tilde{\mathbf{m}} &= \mathbf{H}\mathbf{m} \end{aligned} \quad (3.3)$$

Für die Bestimmung geeigneter Projektionsmatrizen \mathbf{H} und \mathbf{H}' sind mehrere Verfahren entwickelt worden. Neben planaren Projektionen sind auch zylindrische und sphärische Projektionen geeignet [Sei96]. Grundsätzlich wird versucht, die Epipolarbedingung bei einer möglichst geringen Verzerrung der Bilder zu erfüllen. Die Rektifizierung ist bei mehreren bildbasierten Verfahren der erster Schritt zur View-Synthesis. Sie wird nach Seitz [Sei96] daher auch als *Prewarp* bezeichnet.

Die Fundamentalmatrix beschreibt die Geometrie zweier Ansichten. Einige bildbasierte Verfahren der Ansichtengenerierung basieren auf der Geometrie dreier Ansichten [Avi98], [Fau01]. Eine trifokale Ansicht führt zu weiteren Bedingungen, die für die View-Synthesis genutzt werden können. Die dritte Abbildung ist über die Fundamentalmatrizen \mathbf{F}_{12} und \mathbf{F}_{23} bestimmt (siehe Abbildung 3.5):

$$\mathbf{m}'' = \mathbf{F}_{12}\mathbf{m} \times \mathbf{F}_{23}\mathbf{m}' \quad (3.4)$$

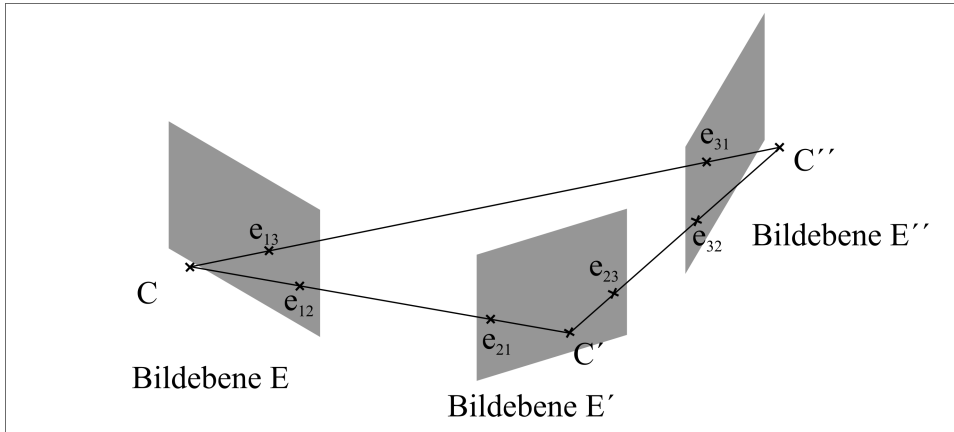


Abbildung 3.5: Trifokale Ansicht; nach [Fau01]

Während die Geometrie zweier Ansichten mit der 3×3 -Fundamentalmatrix beschrieben werden kann, ist für die trifokale Ansicht ein $3 \times 3 \times 3$ -Tensor \mathcal{T}_i^{jk} geeignet. Die grundlegende Trifokal-Bedingung ergibt sich für Bestimmung des Bildes einer Geraden in der dritten Ansicht (siehe Abbildung 3.6):

$$\mathbf{m}^i \mathbf{l}_j'' \mathbf{l}_k'' \mathcal{T}_i^{jk} = 0 \quad (3.5)$$

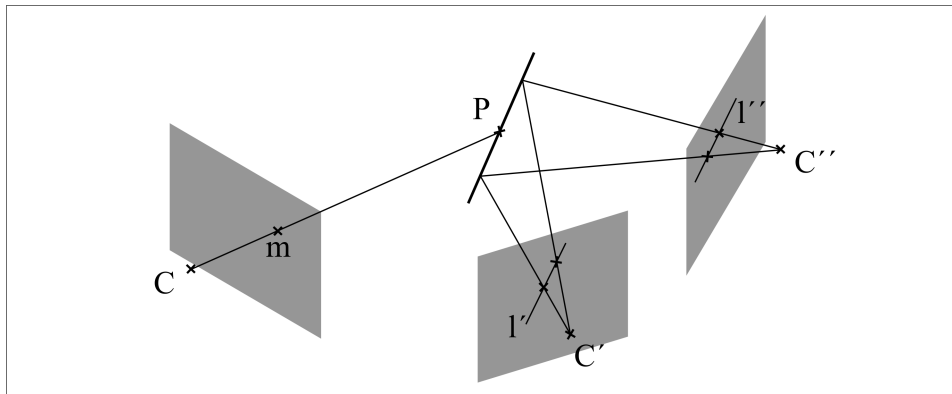


Abbildung 3.6: Bild einer Geraden in der 3. Ansicht; nach [Fau01]

Die Nutzung einer trifokalen Ansicht bietet zwei Vorteile: Die Abbildung eines Punktes auf der dritten Bildebene ist vollständig durch die beiden anderen Abbildungen bestimmbar. Der zweite Vorteil ist die Nutzbarkeit von Korrespondenzen nicht nur zwischen den Abbildungen von Punkten, sondern auch von Geraden (siehe Abbildung 3.6).

3.2.2 Gewinnung von Objektinformationen

Obwohl die bildbasierten Verfahren zur View-Synthesis nicht die Rekonstruktion des 3D-Objektes benötigen, erfordern sie dennoch zusätzliche Informationen über das Objekt und dessen Aufnahme.

Korrespondenzen

Die Zuordnung von Bildpunkten aus verschiedenen Ansichten zu ein und demselben Objektpunkt ist Ausgangspunkt vieler Verfahren der View-Synthesis [Che93], [Ohm97a] [Red00], [Sei96], [Wer95] u. a. Für die Korrespondenzpunktfindung gibt es eine Vielzahl an Verfahren, die im Wesentlichen auf folgenden Methoden beruhen: Merkmalsextraktion, Block-Matching, Bildpunktrekursion, Abschätzung des optischen Flusses, Bayessche Schätzung [Red00].

Disparitätenkarten

Bei parallelen stereoskopischen Aufnahmen weisen Korrespondenzpunktpaare einen horizontalen Versatz (Disparität) in den beiden Bildern auf. Die Disparität ist abhängig vom Abstand des Objektpunktes. Voraussetzung zur Bestimmung der Disparität ist die ermittelte Korrespondenz. Zur Erstellung einer Disparitätenkarte ordnet man den Punkten des linken bzw. des rechten Bildes jeweils einen Disparitätswert zu. Unbekannte Disparitätswerte können durch die Kontinuitäts- und Ordnungseinschränkung, das Kohärenzprinzip oder eine Grob-zu-Fein-Strategie auf Basis der bekannten Werte ermittelt werden [Haa04].

Tiefenkarten

Ein Hilfsmittel für die View-Synthesis sind Tiefenkarten (*disparity maps*). Dabei wird jedem Bildpunkt ein Tiefenwert zugeordnet. In Abbildung 3.7 ist die Ansicht eines Videoobjektes der zugehörigen Tiefenkarte gegenübergestellt. Tiefenkarten können aus den Disparitätswerten bestimmt werden [Ohm97b]. Für ein konvergentes Kamerapaar ergibt

sich die Tiefe Z aus:

$$Z = -F \cdot \frac{B}{d^* + d_0} \quad (3.6)$$

(B Kameraabstand, F Brennweite, d^* gemessene Disparität, d_0 Disparitätswert für den Schnittpunkt der optischen Achsen)

Tiefenkarten werden bei der Ansichtengenerierung für eine perspektivisch korrekte Interpolation von Bildpunkten genutzt.



Abbildung 3.7: Objektansicht mit zugehöriger Tiefenkarte; [Haa04]

Silhouette

Ein dreidimensionales Objekt wird in jeder aufgenommenen Ansicht durch eine Silhouette berandet. Die Silhouette entspricht in den einzelnen Aufnahmen der so genannten Maske, deren Codierung als Alpha-Werte bereits standardisiert ist [ISO98]. Die notwendige Segmentierung kann für Blue-Screen-Aufnahmen recht einfach mittels Chroma-Keying erfolgen. Die Silhouetten-Information ist einerseits eine Bildinformation, wird aber andererseits bei einigen Verfahren auch zur Ansichtengenerierung genutzt [Pol98].

Image-based Visual Hull

Die Visual Hull ist definiert als die maximale geometrische Annäherung an ein 3D-Objekt, welche durch die Bildung der Schnittmenge aller Silhouettenkörper erzielt werden kann [Lau94] (entspricht einer *volume intersection*; siehe Abbildung 3.8). Zur Bestimmung der Visual Hull wurden bildbasierte Verfahren entwickelt, welche die Eigenschaften der Epipolargeometrie nutzen [Mat00].

Layered Depth Images (LDI)

Die LDI-Repräsentation ist für das Image-Based-Rendering (IBR) entwickelt worden und dient zur perspektivisch korrekten Darstellung [Sha98]. Diese Repräsentation ordnet die Bildpunkte inkl. Tiefeninformation schichtenweise in einer zweidimensionalen Matrix an.

3.2.3 View-Synthesis

Für die Generierung neuer Ansichten sind eine Vielzahl an bildbasierten Verfahren in Entwicklung. Im Folgenden werden die wichtigsten Verfahren erläutert.

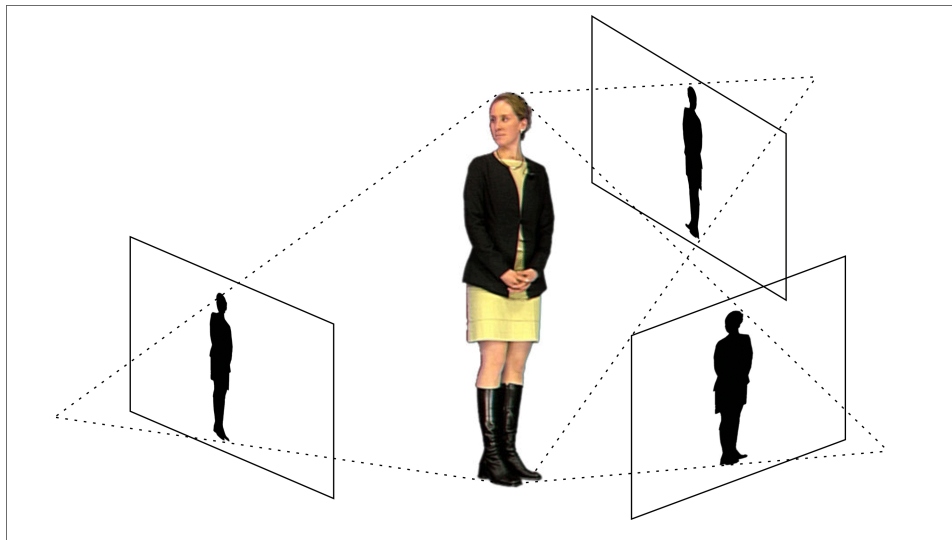


Abbildung 3.8: Generierung der Visual Hull aus den Silhouetten

Morphing

Ein grundlegender Ansatz zur Generierung einer neuen Ansicht ist die Interpolation zwischen zwei bestehenden Ansichten. Man geht hierbei davon aus, dass die beiden bekannten Ansichten den Anfang und das Ende eines Umwandlungsprozesses darstellen. Mittels einer analytischen Beschreibung dieser Metamorphose können beliebige Stationen, d. h. die interpolierten Ansichten generiert werden. Da die Ansichten durch Bildpunktmatrizen repräsentiert werden, muss die Metamorphose der Bildpunkte beschrieben werden. Dazu müssen die Korrespondenzen in den beiden Ansichten bekannt sein. Chen und Williams nutzen für ihr Verfahren die Kameraparameter zur automatischen Korrespondenzdeterminierung [Che93], wofür eine fixe Kalibrierung notwendig ist. Die Metamorphose für die Bildpunkte wird durch Vektoren beschrieben, sodass Zwischenansichten interpoliert werden können. Diese lineare Mittelung führt jedoch nicht immer zu perspektivisch korrekten Ergebnissen.

Pollard und Hayes nutzen die Kanten der Silhouetten als korrespondierende Merkmale [Pol98]. Es erfolgt eine lineare Interpolation der Baryzentren zwischen drei Ansichten. Das Verfahren arbeitet korrekt, wenn die Silhouetten dieselben Objektbereiche repräsentieren und keine Okklusionen vorhanden sind.

Seitz und Dyer entwickelten ein Morphing-Verfahren, welches auf der Geometrie der mehrfachen Ansicht¹ beruht [Sei96], [Sei97]. Dadurch ist eine perspektivisch korrekte Interpolation möglich. Das Verfahren erfolgt in drei Schritten: *Prewarp* (Rektifizierung), *Morph* (z. B. lineare Interpolation von Korrespondenzen), *Postwarp* (Derektifizierung). Die Rektifizierung ist in Unterabschnitt 3.2.1 erläutert worden und hat das Ziel, dass die Korrespondenzpunktpaare jeweils in den gleichen Bildzeilen der rektifizierten Bilder liegen. Die Bildebenen E_0 und E_1 in Abbildung 3.9 entsprechen den Ausgangsbildern und die Bildebenen \tilde{E}_0 und \tilde{E}_1 den zugehörigen rektifizierten Ansichten. Dadurch kann der Morphing-Prozess eindimensional in den Bildzeilen erfolgen. Dies hat den wesentlichen Vorteil, dass Korrespondenzen nur innerhalb der Bildzeile zu suchen sind. Mit Hilfe der sich daraus ergebenden Disparitäten sind durch lineare Interpolation die Disparitäten für \tilde{E}_S zu bestimmen (siehe Abbildung 3.10). Die Ansicht \tilde{E}_S hat dasselbe optische Zentrum

¹nach [Fau93]

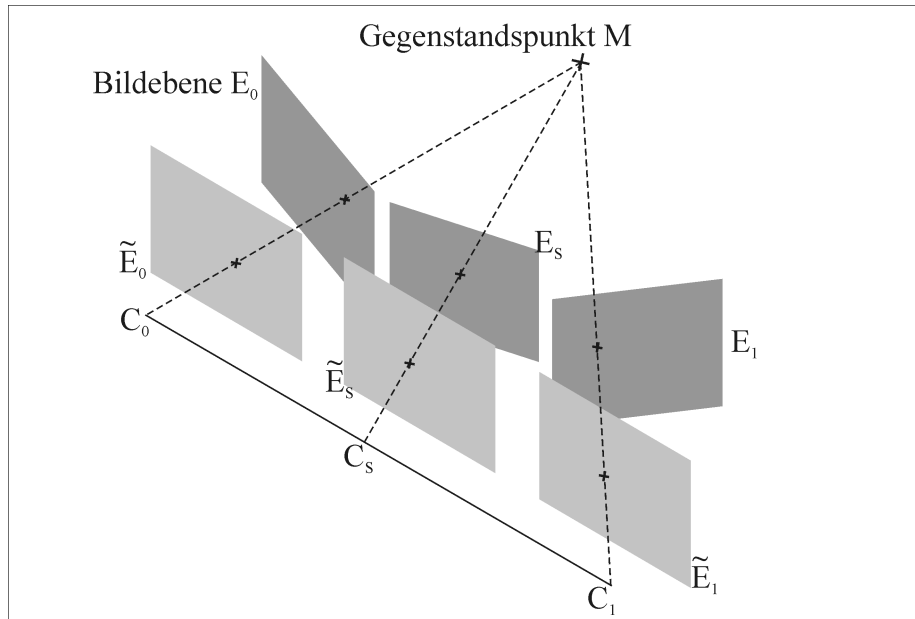


Abbildung 3.9: Morphing nach [Sei96] in drei Schritten

wie die gewünschte Ansicht E_S . Die Berechnung von E_S erfolgt im dritten Schritt. Die Derektifizierung projiziert schließlich die interpolierte Ansicht auf die gewünschte Bildebene.

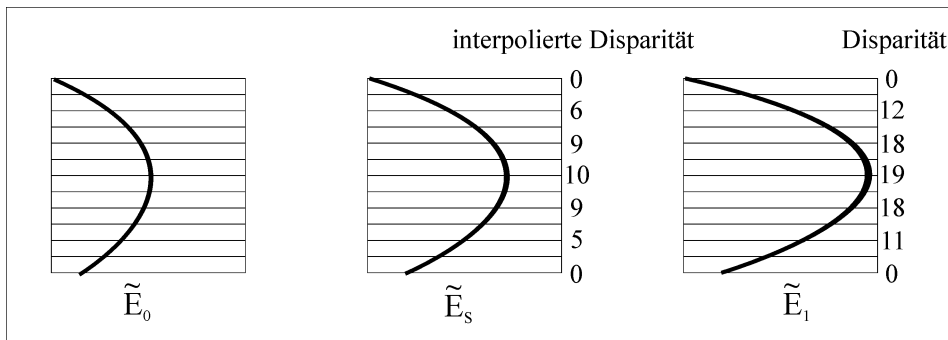


Abbildung 3.10: Lineares Morphing von Bildpunkten

Das Verfahren nach Seitz und Dyer ist geeignet, perspektivisch korrekte Ansichten zu erzeugen. Dies setzt jedoch eine exakt ermittelte Epipolargeometrie, exakt ermittelte Korrespondenzen und das Nichtvorhandensein von Okklusionen voraus.

Incomplete 3D

Ohm und Müller entwickelten ein Verfahren zur Erzeugung von 3D-Videoobjekten, welches die Vorteile der disparitätsbasierten Ansichteninterpolation mit den Vorteilen der 3D-Texturierung vereinigt [Ohm97a]. Da es ansatzweise die Prinzipien der 3D-Computergrafik nutzt, wird es *Incomplete 3D (I3D)* genannt. Außerdem wurde im Gegensatz zu anderen Verfahren versucht, es eng mit einer effektiven Codierung zu verknüpfen. Es erfolgt eine objektbasierte Codierung von Textur-, Tiefen- und Hilfsdaten gemäß MPEG-4 Video.

Für die Generierung einer I3D-Textur werden in den verfügbaren Ansichten so genannte *Areas of Interest (AOI)* selektiert und zu einer Textur vereinigt, die alle sichtbaren Teile

des Objektes enthält. Es werden mit Hilfe der zu ermittelnden Disparitäten immer die Bereiche ausgewählt, welche der jeweiligen Kamera am nächsten sind. Die AOI werden zur *primary surface* zusammengefasst. Zur Vermeidung von Artefakten (z. B. auf Grund von Okklusionen) wird eine *secondary surface* aus den jeweiligen Komplementen der AOI gebildet. Beide Texturen werden mittels spezieller Wichtungsfunktionen (abhängig von Ort, Disparität etc.) zu einer *final surface* zusammengefasst. In Abbildung 3.11 sind die Schritte bis zur I3D-Repräsentation dargestellt.

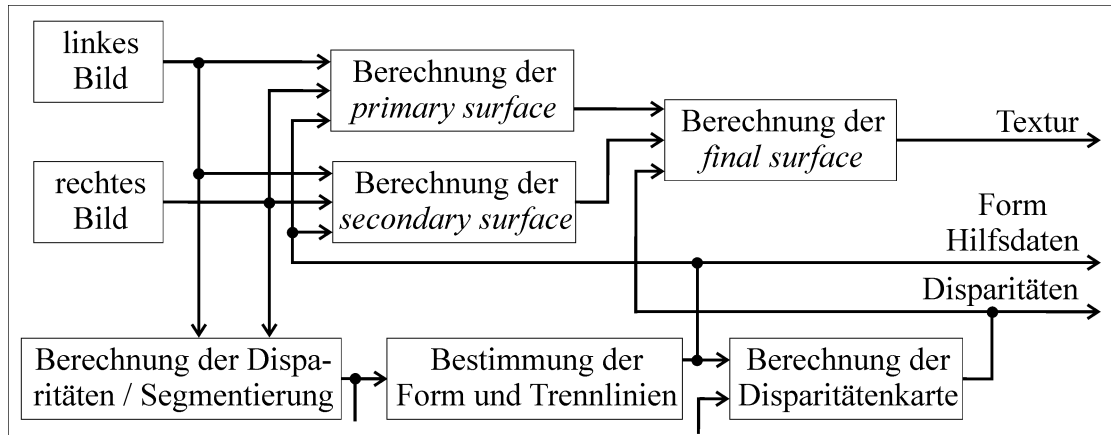


Abbildung 3.11: Bildung der I3D-Repräsentation; nach [Ohm99a]

Die View-Synthesis auf Basis der I3D-Repräsentation ist auf Grund der erfolgten Vorverarbeitung vergleichsweise wenig aufwändig. Entsprechend dem gewählten Blickpunkt ist die I3D-Textur disparitätsgesteuert auf die Bildebene zu projizieren. Dies entspricht einer eindimensionalen Kontraktion der Bildpunkte (siehe Abbildung 3.12). Naturgemäß verbleiben nach der Projektion einige Fehlstellen, z. B. auf Grund diskontinuierlicher Disparitätenkarten. Diese Fehlstellen werden mittels Interpolation geschlossen [Ohm99a].

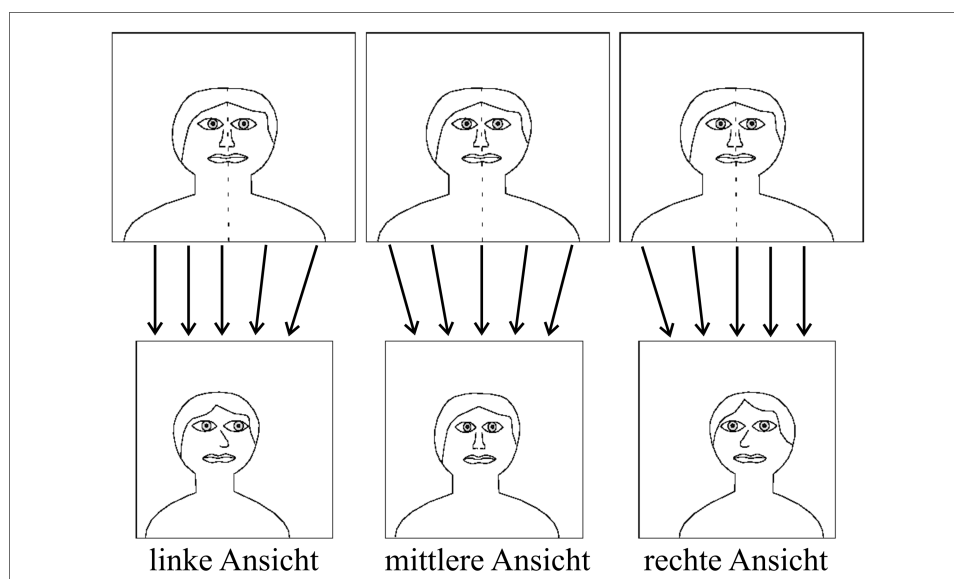


Abbildung 3.12: View-Synthesis der I3D-Repräsentation; nach [Ohm99a]

Light-Field-Rendering, Ray-Space-Rendering

Das Rendering des so genannten Light-Field ist eine Methode des Image-Based-Rendering aus dem Bereich der Computergrafik und -animation. Beim Image-Based-Rendering werden Bilder anstatt von Modellen als Quelle benutzt. Diese Bilder werden als „Scheiben“ des mit Lichtstrahlen gefüllten Raumes aufgefasst. Ausgangspunkt zur Beschreibung dieses Raumes ist die plenoptische Funktion, welche sämtliche visuelle Information von einem gegebenen Betrachterstandpunkt beschreibt. Diese Funktion $p = P(\Theta, \Phi, \lambda, V_X, V_Y, V_Z, t)$ ist von sieben Variablen abhängig: Θ, Φ : Richtung; λ : Wellenlänge des Lichtes; V_X, V_Y, V_Z : Punkt im Raum; t : Zeit [McMil95]. Beim Light-Field-Rendering reduziert sich die Funktion auf vier Dimensionen, da eine statische Szene mit begrenzenden Flächen betrachtet wird und statt der Wellenlänge λ die Grundfarben verwendet werden. Jeder Lichtstrahl kann durch die Schnittpunkte mit zwei gegebenen Flächen beschrieben werden. Levroy und Hanrahan nennen diese Repräsentation *Light Slab* (siehe Abbildung 3.13) [Lev96]. Ebenfalls in Abbildung 3.13 ist ein Beispiel für das Feld der Lichtstrahlen zwischen zwei Flächen dargestellt. Diese beiden Flächen sind entsprechend dem zu erfassenden Strahlenfeld zu wählen. Eine der beiden Flächen kann hierfür im Unendlichen liegen.

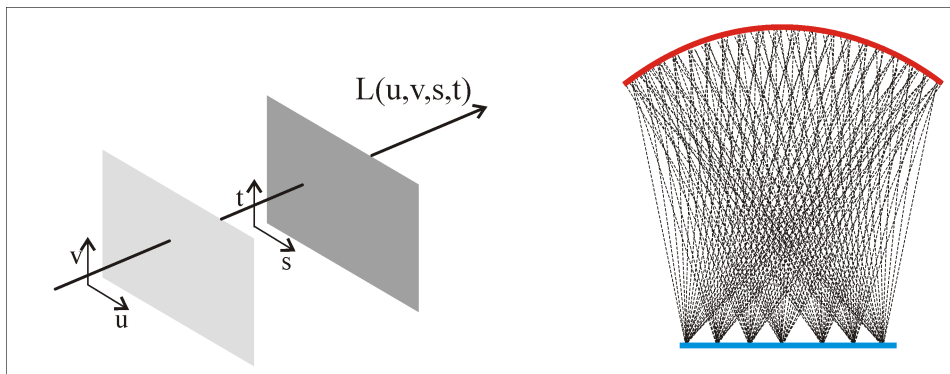


Abbildung 3.13: Light-Field-Rendering: Light Slab (links), Feld der Lichtstrahlen zwischen 2 Flächen (rechts)

Eine weitere Repräsentation der Lichtstrahlen im Raum ist der *Ray-Space* [Fuj96], [Nae00], [Smo03]. Diese Repräsentationsform wurde ursprünglich als Datenformat für die dreidimensionale visuelle Kommunikation entwickelt. Der Ray-Space enthält alle visuellen Informationen aus jeder beliebigen Blickrichtung. Jedes Bild des Raumes wird dabei als ein 2D-Unterraum des 4D-Ray-Space aufgefasst. Die Ray-Space-Repräsentation unterscheidet sich vom Light-Field lediglich in der Beschreibung der Lichtstrahlen. Im Ray-Space wird jeder Strahl ausgehend von einem Punkt im Raum als $f(x, y, \theta, \phi)$ in kartesischen Koordinaten oder als $f(\xi, \zeta, \theta, \phi)$ in sphärischen Koordinaten angegeben.

3.3 Modellbasierte Verfahren

Im Gegensatz zu bildbasierten Verfahren wird bei modellbasierten Verfahren versucht, ein zeitveränderliches 3D-Modell des aufgenommenen Objektes zu rekonstruieren. Diese Methode führt zu den Grundprinzipien der 3D-Computergrafik und -animation. Nach der Rekonstruktion kann dieses 3D-Modell nämlich wie jedes andere 3D-Computergrafikmodell im Rendering-Prozess verwendet werden. Dadurch ergibt sich gegenüber bildbasierten Verfahren der Vorteil, dass das 3D-Videoobjekt derselben Beleuchtungssituation unterliegt

sowie Schatten werfen und empfangen kann.

3.3.1 Aufnahme

Die Aufnahme erfolgt ähnlich wie bei den bildbasierten Verfahren mittels mehrerer Kameras (Multi-View), die um das aufzunehmende Objekt herum positioniert sind. Zur Unterstützung der Gewinnung der räumlichen Information können weitere Aufnahmesysteme, wie Tiefenkameras² oder LIDAR³ verwendet werden.

3.3.2 Extraktion eines 3D-Modells

Die Extraktion von räumlichen Informationen aus Bildern ist eine der zentralen Fragestellungen der Computervision. Hierfür ist eine Vielzahl von Verfahren entwickelt worden. Ein Großteil der Verfahren ist jedoch nur für statische Objekte geeignet, sodass sie für 3D-Videoobjekte nicht genutzt werden können. Für die Extraktion eines 3D-Modells werden verschiedene Klassen an Verfahren genutzt [Eis99]:

- Verfahren, welche die Oberfläche des 3D-Objektes aus den gewonnenen Tiefenkarten bestimmen (siehe unter 3.2.2: Tiefenkarten)
- Verfahren auf der Basis einer Volume-Intersection (siehe unter 3.2.2: Image-based Visual Hull). Hierfür wird meist die Silhouetten-Information genutzt [Mat04], [Lau94], [Wür02], [Che00], [Moe97]. Dies setzt selbstverständlich eine ausreichend genaue Ermittlung der Silhouette, d. h. Segmentierung voraus.
- Einige Verfahren nutzen die Tiefeninformationen und die Ergebnisse der Volume-Intersection [Eis99], [Sei99]. Diese Verfahren erlauben eine gleichzeitige Texturierung der Oberfläche (siehe unten: Voxel-Colouring).

Für die Repräsentation des 3D-Modells gibt es verschiedene Methoden, die meist eng an den Extraktionsprozess gebunden sind:

- *dynamische 3D-Gittermodelle*: Ebenso wie in der konventionellen Computergrafik finden auch dynamische 3D-Gittermodelle Verwendung [Isi02].
- *Voxel*: Viele Verfahren nutzen die Voxel-Repräsentation, welche gleichzeitig zur Beschreibung der farblichen Eigenschaften dienen kann [Sei99], [Mül04], [Mik01], [Eis99]. Für eine vereinfachte (schnelle) Verarbeitung bietet sich eine hierarchische Octree-Struktur zur Organisation der Voxel an [Bær98], [Mül04].
- *3D-Videoartikel/-fragmente*: Weitere Repräsentationsformen sind 3D-Videoartikel und 3D-Videofragmente [Wür02], [Wür04]. Diese dynamischen Repräsentationen erlauben eine effektive Codierung und ermöglichen ein hardwareunterstütztes Rendering.
- *Nutzung mehrerer Repräsentationsformen*: z. B. eine Überführung eines Voxel-Modells in ein Drahtgittermodell [Mül04]

²z. B. Z-CamTM

³light detection and ranging

3.3.3 Texturierung

Neben einer Beschreibung der dreidimensionalen Objektgeometrie ist die Beschreibung des Aussehens, d. h. die Farbe der Oberfläche, notwendig. Für diese Texturierung gibt es verschiedene Ansätze [Mül04]:

- *statisch*: Eine statische Texturierung führt oft zu unzureichenden Ergebnissen, da das Aussehen der Oberfläche nicht blickpunktabhängig ist (\rightarrow *painted shoebox effect*).
- *blickpunktabhängig (view dependent)*: Es wird eine blickpunktabhängige Textur verwendet, welche das Ergebnis einer gewichteten Mittelung der sich aus den Aufnahmen ergebenden Texturen ist. Hierfür werden die für den neuen Blickpunkt relevanten Ansichten entsprechend der Sichtbarkeit auf der Objektoberfläche gemischt [Deb96] (siehe Abbildung 3.14).
- *Light-Field-Mapping*: Für die einzelnen Texturen wird eine Light-Field-Repräsentation genutzt (siehe 3.2.3: Light-Field-Rendering). Hierfür ist die allgemeine Light-Field-Beschreibung $f(r, s, \theta, \phi)$ mittels Wichtung den Vertices der Objektoberfläche anzupassen [Che02].
- *Unstructured Lumigraph*: Es erfolgt ein IBR auf Basis von beliebigen Aufnahmen (beliebiger Blickpunkt/-richtung etc.). Mittels spezieller Blending-Methoden werden in Abhängigkeit von Blickwinkel, Seitenverhältnis und Auflösung/Entfernung die Ansichten zu einer äquivalenten Textur gemischt [Bue01]. Die so entstehenden Lumigraph-Repräsentationen sind nichts anderes als Light-Field-Beschreibungen.

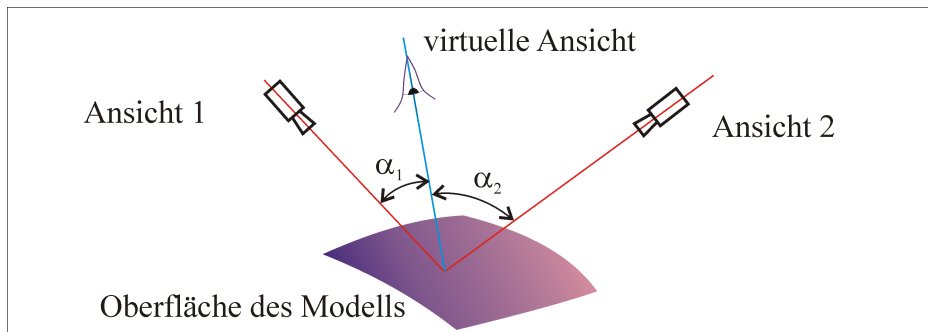


Abbildung 3.14: Bestimmung einer blickpunktabhängigen Textur; nach [Deb96]

3.3.4 Voxel-Colouring

Einige Verfahren trennen die Extraktion des 3D-Modells und die Texturierung nicht grundsätzlich voneinander. Insbesondere gilt dies für Voxel-basierte Verfahren wie das Voxel-Colouring [Sei99]. Da Voxel gleichzeitig die Form und die Farbe des Objektes beschreiben, ist die Bestimmung der Farbe der Objektvoxel eine vollständige Rekonstruktion des Objektes. In Abbildung 3.15 ist das Grundprinzip des Voxel-Colouring nach [Sei99] dargestellt, welches Voxel für Voxel durchgeführt wird. Nach der Auswahl eines Voxels werden dessen Projektionsorte in den Ansichten $E_1 \dots E_n$ bestimmt. Anschließend wird untersucht, ob die Bilder des Voxels übereinstimmen (*consistency check*). Hierfür gilt es,

ein geeignetes Abstandsmaß inkl. Schwelle festzulegen. Eine weitere Schwierigkeit ist die Behandlung von Okklusionen (siehe E_3). Wird die Konsistenz für den Voxel festgestellt, wird er dem Objekt zugeordnet und erhält entsprechend den relevanten Abbildungen $E_x \dots E_y$ einen Farbwert.

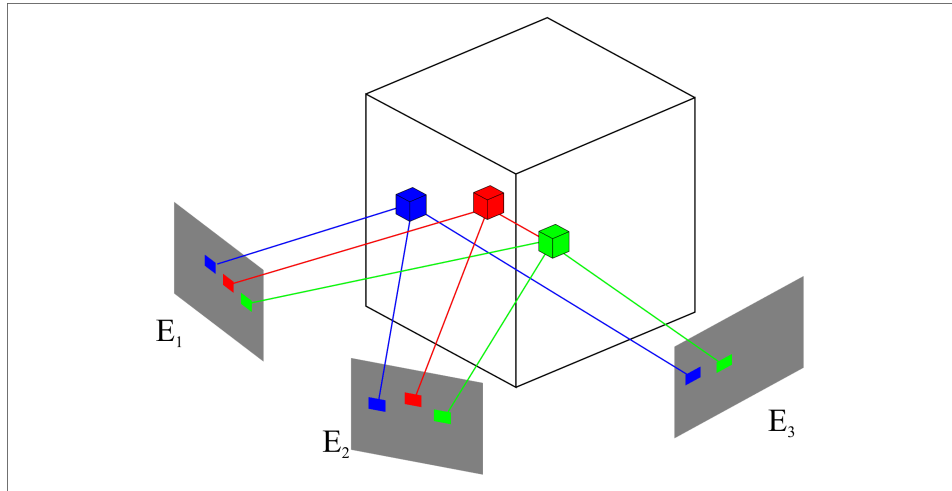


Abbildung 3.15: Voxel-Colouring

3.3.5 Relief-Texture-Mapping

Ein weiterer Ansatz zur Repräsentation von 3D-Videoobjekten basiert auf dem Relief-Texture-Mapping [deOli00]. Das 3D-Modell ist dabei ein einfacher umhüllender Quader. Dieser hat sechs Texturen, welche sowohl Farb- als auch Tiefeninformationen beinhalten. Nach dem Rendering-Prozess ist nur noch das beliebig geformte 3D-Modell zu sehen. Ein Vorteil dieses Verfahrens ist die Unterstützung des Rendering-Prozesses durch moderne PC-Grafikhardware. Spezielle Verfahren zur Bestimmung dieser Texturen aus multiskopischen Aufnahmen für 3D-Videoobjekte sind noch nicht bekannt.

3.4 Überblick über die Verfahren

In den beiden vorherigen Abschnitten wurden Verfahren zur Erzeugung von 3D-Videoobjekten dargestellt. Es wurden einerseits bild- und andererseits modellbasierte Verfahren erläutert. Bei der Betrachtung der Verfahren wurde jedoch deutlich, dass diese einfache Einteilung schnell an ihre Grenzen stößt. So wird einerseits bei den bildbasierten Verfahren versucht, Informationen über die Objektgeometrie einzubeziehen. Andererseits wird bei den modellbasierten Verfahren versucht, Methoden des IBR einzubeziehen. Die Grenze zwischen bild- und modellbasierten Verfahren ist fließend. Eine Einteilung, welche diesem Umstand Rechnung trägt, wurde von Pollefeys in [Pol00] vorgestellt (siehe Abbildung 3.16). Sämtliche Verfahren lassen sich auf einer Achse einordnen, je nachdem ob das Verfahren einen mehr bildbasierten oder einen mehr modellbasierten Charakter hat.

Die Darstellung der verschiedenen Verfahren zur Erzeugung von 3D-Videoobjekten führt abschließend zu folgende Erkenntnissen:

- Der Aufwand für die Generierung eines 3D-Videoobjektes ist mit den derzeit verfügbaren Mitteln immens. Dies beginnt bei der Aufnahme, da möglichst viele Ansichten

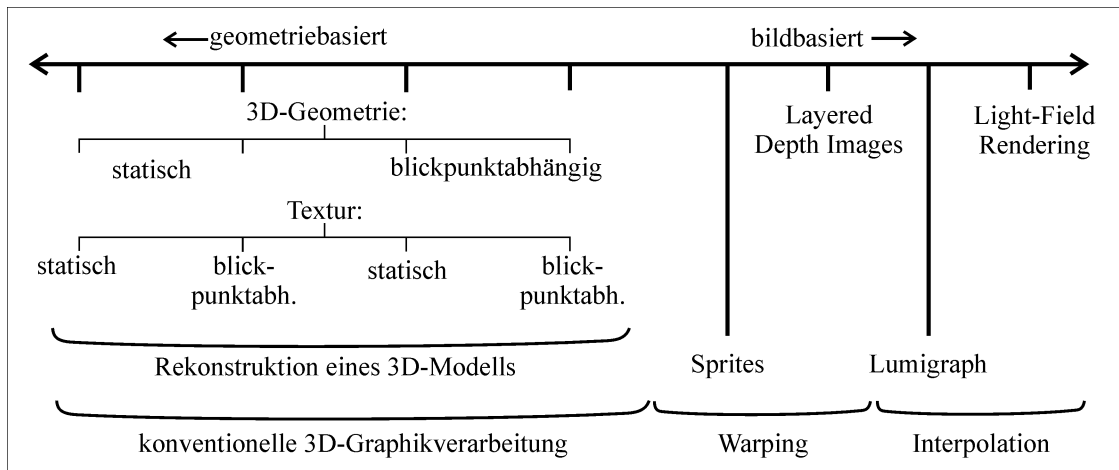


Abbildung 3.16: Verfahren zur Generierung von 3D-Videoobjekten; nach [Pol00]

benötigt werden.

- Die Verfahren weisen noch keine hohe Reife auf.
- Auf Grund der Unterschiedlichkeit der Anwendungen (z. B. immersive Telekonferenz, Virtual Reality) bestehen unterschiedliche Anforderungen, welche durch die verschiedenen Verfahren unterschiedlich erfüllt werden. Daher ist davon auszugehen, dass sich künftig nicht nur ein Verfahren etablieren wird.
- Die aktuellen Entwicklungen zeigen, dass die Kombination aus bild- und modellbasierten Prinzipien am vielversprechendsten ist.

Für die Generierung eines 3D-Videoobjektes lassen sich folgende Kriterien erkennen:

- *Aufwand für die Aufnahme:* z. B. Anzahl der Kameras und Hilfseinrichtungen, Segmentierungshilfen (Blauwand)
- *entstehende Datenmenge:* Anzahl der synchronen Sequenzen (bestehend jeweils aus N Texturen, N Tiefenkarten etc.)
- *Aufwand für die Vorverarbeitung:* z. B. Segmentierung, Korrespondenzsuche (automatisch/manuell), Rektifizierung, Modellrekonstruktion, Texturierung etc.
- *Adaptierbarkeit:* an bestehende Aufnahmetechniken/-studios, an existierende Codierverfahren, an gängige (hardwareunterstützte) Rendering-Verfahren
- *View-Synthesis:* Aufwand für Warping, Rendering etc.; Realzeit-Fähigkeit
- *Nutzungsmöglichkeiten:* Größe des zulässigen Navigationsbereiches um das Objekt; zulässige Nähe
- *Qualität*

Selbstverständlich sind diese Kriterien miteinander verknüpft. Ihre Gegenläufigkeit ist in Abbildung 3.17 skizziert.

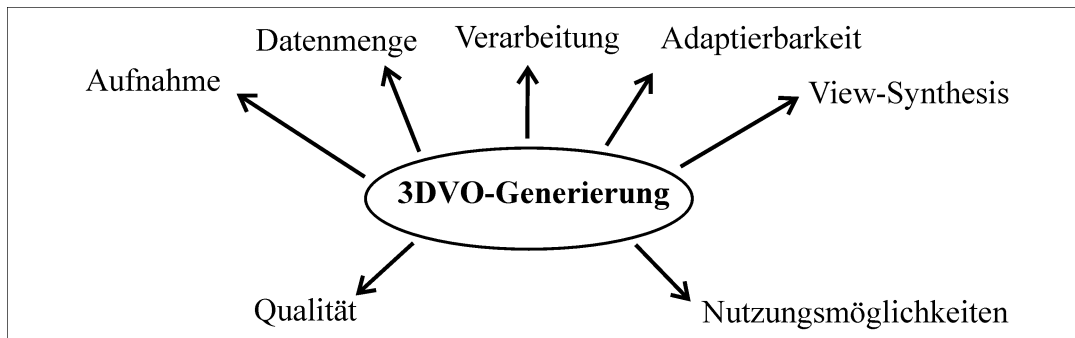


Abbildung 3.17: Kriterien für die Generierung von 3D-Videoobjekten

3.5 Modell der 3D-Videoobjektgenerierung

Da eine Bewertung von 3D-Videoobjekten erfolgen soll, welche unabhängig von deren Generierungsverfahren ist (siehe Abschnitt 1.2), wird ein allgemeingültiges Modell zur Beschreibung der 3DVO-Generierung benötigt. Es ist erforderlich, um einerseits einheitliche Termini verwenden zu können und um andererseits die verschiedenen Prozesse vergleichen zu können. Hierfür wurden zunächst Ausgangspunkt und Ziel eines 3D-Videoobjektes definiert (siehe Abschnitt 3.1). Alle Verfahren lassen sich in eine zweistufige Verarbeitungskette unterteilen. Der Ansichtengenerierung (View-Synthesis, Rendering) geht immer eine Vorverarbeitung voraus. In der Vorverarbeitung werden aus den aufgenommenen Daten (*primäre Repräsentation*) die für die Ansichtengenerierung erforderlichen Daten (*sekundäre Repräsentation*) erzeugt. In Abbildung 3.18 ist das Modell der 3DVO-Generierung dargestellt. Der Terminus *3D-Videoobjekt* bezieht sich hierbei auf das Ergebnis der Kette, d. h. die generierten Ansichten. Mitunter wird der Begriff entsprechend den modellbasierten Verfahren für die 3D-Modelle (inkl. Textur) verwendet. Diese Begriffsverwendung ist nicht übertragbar auf bildbasierte Verfahren. Daher wird der Begriff im Folgenden für das Ergebnis des 3DVO-Prozesses verwendet.

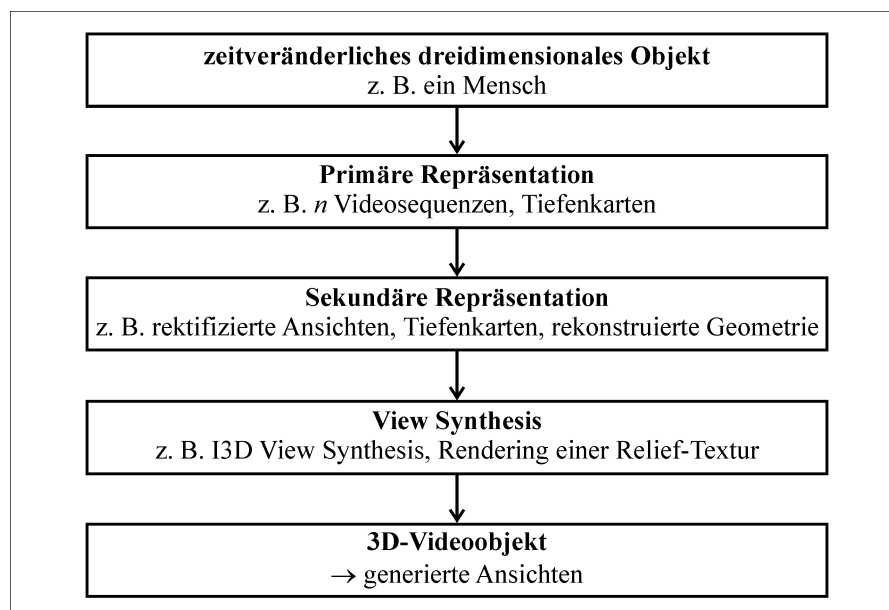


Abbildung 3.18: Modell der Generierung von 3D-Videoobjekten; nach [Rit03], [Rit04]

4 Modell der Qualität von 3D-Videoobjekten

4.1 Zielstellung für die Modellbildung

Die objektive Beurteilung der Qualität von 3D-Videoobjekten erfordert ein Modell, mit dem man diese Qualität beschreiben kann. Hierfür sind zunächst die Ziele dieser Qualitätsbewertung festzulegen:

- Es sollen 3D-Videoobjekte, welche mittels unterschiedlicher Verfahren generiert wurden, bewertet werden können (siehe Abschnitt 1.2).
- Es sind die spezifischen Qualitätsmerkmale von 3D-Videoobjekten in die Bewertung einzubeziehen.
- Die Methodik der Qualitätsbewertung soll für künftige 3DVO-Verfahren anwendbar sein.
- Das Modell soll durch subjektive Tests verifizierbar sein.

Diese Ziele unterstreichen den umfassenden Ansatz zur Qualitätsbewertung von 3D-Videoobjekten auf Basis der in Abschnitt 3.1 getroffenen Definition. Ebenso ist eine Abgrenzung vorzunehmen, welche Aspekte der Qualitätsbewertung mit diesem Modell nicht behandelt werden sollen:

- eine Bewertung einer primären Repräsentation
- eine Bewertung einer sekundären Repräsentation
- eine Bewertung eines einzelnen Abschnittes der 3DVO-Verarbeitung (z. B. das Rendering des 3D-Modells)
- eine Beurteilung des zulässigen Navigationsbereiches um das Objekt
- die Bewertung der Inklusion

Das Modell der 3DVO-Generierung (siehe Abbildung 3.18) verdeutlicht, dass ein Vergleich von verschiedenen 3DVO-Verfahren auf der Basis der Repräsentationen nicht möglich ist. Es wird nicht gelingen, beispielsweise rektifizierte Ansichten mit einem 3D-Modell in deren Qualität zu vergleichen. Selbstverständlich kann es ebenso nicht das Ziel sein, die unterschiedlichen Verarbeitungsschritte zu bewerten. Hierfür kann es jeweils nur spezielle Verfahren geben, die beispielsweise unterschiedliche Rendering-Verfahren für ein modellbasiertes 3DVO bewerten.

Eine Beurteilung des zulässigen Navigationsbereiches um das Objekt soll im Rahmen dieser Qualitätsbewertung nicht erfolgen. Der zulässige Navigationsbereich wird durch die Anwendung definiert. Wenn ein 3D-Videoobjekt den geforderten Navigationsbereich nicht zulässt, ist es für die jeweilige Anwendung ungeeignet und eine Qualitätsbewertung erübrigt sich.

Für den Nutzer einer audiovisuellen Anwendung mit 3D-Videoobjekten wird die Inklusion des Objektes den subjektiven Qualitätseindruck beeinflussen. Dieses Empfinden ist u. a. davon abhängig, inwieweit die Beleuchtung von Objekt und Szene übereinstimmen und ob die Tiefenschärfe der Szene und die Tiefenschärfe des Objektes harmonisieren. Dies sind selbstverständlich Qualitätsmerkmale der gesamten Szene. Daher kann die Inklusion für die Bewertung einzelner 3D-Videoobjekte nicht berücksichtigt werden.

Die Untersuchung der Verfahren zur Qualitätsbewertung von konventionellem Video (siehe Kapitel 2) und der 3DVO-Verfahren (siehe Kapitel 3) sowie die Bestimmung der Ziele der Qualitätsbewertung führen zu folgenden Schlüssen:

- *Conclusio 4.1*: Der Vergleich von 3D-Videoobjekten kann nur auf der Ebene der generierten Ansichten erfolgen.
- *Conclusio 4.2*: Zur Qualitätsbewertung ist eine Regression der Qualitätsmerkmale vorzunehmen.
- *Conclusio 4.3*: Die Qualitätsbewertung hat sich auf einen typischen bzw. auf einen kritischen Anwendungsfall zu beziehen.

Die erste *Conclusio* ergibt sich unmittelbar aus dem Hauptziel der Bewertung und dem Modell der 3DVO-Generierung.

Für die zweite *Conclusio* bestand die Fragestellung, ob die Qualitätsbewertung auf Basis von ermittelten Qualitätsmerkmalen oder mittels der Nachbildung des HVS erfolgen soll. Die derzeit verfügbaren HVS-Nachbildungen berücksichtigen noch nicht die Wahrnehmung von 3DVO-Merkmalen¹ durch den Visus. Diese Nachbildung erfordert ein fundiertes Wissen über die Phänomene des Visus, welche bei der Nutzung/Betrachtung eines 3D-Videoobjektes auftreten. Daher wird eine Regression der Qualitätsmerkmale vorgenommen.

Die dritte *Conclusio* ist der Tatsache geschuldet, dass ein 3D-Videoobjekt zu einem gegebenen Zeitpunkt T (unendlich) viele Zustände haben kann, da dieser Zustand vom Blickpunkt abhängig ist. Dies ist ein grundlegender Unterschied zu konventionellem Video (siehe Unterabschnitt 2.4.5). Daher stellt sich die Frage, welcher Zustand bzw. welche Zustände für die Bewertung genutzt werden sollen. Diese Frage ist teilweise vergleichbar mit der Bewertung beispielsweise eines neuen Videoencoders. Hierfür ist es ebenso wenig möglich, diesen mit jedem potenziell auftretenden Bildmaterial zu testen. Daher verwendet man Testszenen, welche das typische oder das kritische Material repräsentieren. Diese Methodik kann man auf das Problem der vielen Zustände von 3DVO übertragen, d. h. man verwendet für die Bewertung eine typische bzw. kritische Nutzung des 3DVO. Das kann ein typischer Umlauf um das Objekt oder eine Bewegung auf das Objekt zu sein.

4.2 Bisherige Arbeiten zur Qualitätsbewertung von dreidimensionalem Video

Die bisherigen Arbeiten zur Qualitätsbewertung von 3D-Video beziehen sich auf konkrete Verfahren oder auf bestimmte Verarbeitungsschritte. Diese Arbeiten spiegeln die Unterschiedlichkeit der mit Schlagwort *3D-Video* assoziierten Technologien wider. Beispielsweise wurde die Qualitätsbewertung von datenkomprimierten stereoskopischen Fernseh Bildern

¹z. B. Perspektivenfehler; siehe Abschnitt 4.4

untersucht [Kay96]. In anderen Arbeiten erfolgte die Bewertung von Rendering- und Animationsverfahren [Mys01], [Lue01]. Diese lassen sich prinzipiell auch auf modellbasierte 3D-Videoobjekte anwenden. Auch die Qualitätsbewertung von bildbasierten Repräsentationen von 3D-Videoobjekten wird untersucht. Im Zusammenhang mit der Entwicklung von Codierverfahren für Bildsequenzen mit Tiefenbildern muss die zu erwartende Qualität abschätzbar sein [Gre03].

In der MPEG 3DAV-Arbeitsgruppe wird u. a. versucht, modellbasierte Szenenrekonstruktionsmethoden in ihren Ergebnissen zu vergleichen [ISO03b]. Hierfür werden Testsets von multiskopischen Ansichten (synthetisch, natürlich) vorgegeben. Eine der gegebenen Ansichten wird während des Testes ausgelassen. Genau diese Ansicht ist durch das zu untersuchende Verfahren zu rekonstruieren. Zur Qualitätsbewertung wird die ausgelassene mit der synthetisierten Ansicht verglichen. Für die objektive Bewertung der Objekttextur werden das PSNR und dessen Varianz berechnet. Als objektives Maß für die Bewertung der Silhouette dient der mittlere Anteil an korrekt zugeordneten Pixeln.

4.3 Referenzen für die 3DVO-Qualitätsbewertung

4.3.1 Vergleich von 3D-Videoobjekten

Eine grundlegende Frage für die Qualitätsbewertung von 3D-Videoobjekten ist der Bezug zu einer Referenz (*ground truth*). Diese Frage führt zu folgenden Teilfragen:

1. Ist eine Bewertung von 3D-Videoobjekten ohne Referenz (im Sinne von NR²) möglich?
2. Was ist die Referenz für ein 3D-Videoobjekt?

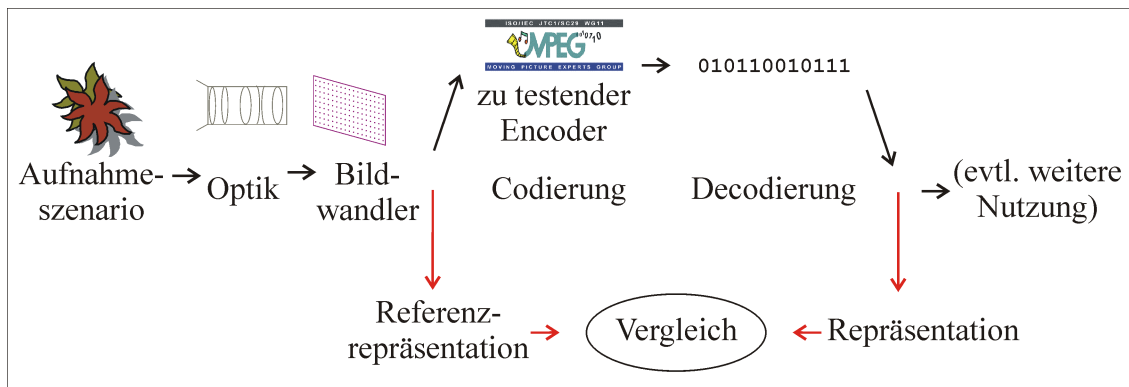


Abbildung 4.1: Beispiel für eine 2D-Videoqualitätsbewertung

Die erste Frage ist für das hier gestellte Ziel der Qualitätsbewertung zu verneinen. Da unterschiedlichst generierte 3D-Videoobjekte verglichen werden sollen, ist eine Vielzahl an Qualitätsmerkmalen zu erfassen. Die Einschränkungen, welche NR-Verfahren zur Messung der Qualität erfordern³, können hier nicht vorgenommen werden. Die zweite Frage wird naturgemäß immer im Zusammenhang mit 3D-Video diskutiert und meist damit beantwortet, dass es keine Referenz gibt. Dies ist richtig, wenn man diese Frage im gleichen Sinn wie

²no reference; siehe Unterabschnitt 2.4.5

³z. B. die Messung von Blockausfällen für MPEG-2-Sequenzen

für die 2D-Videoqualitätsbeurteilung beantwortet. Bei der 2D-Videoqualitätsbeurteilung gibt es fast immer eine eindeutig zuordenbare Repräsentation mit gleichem Aufbau, welche zum Vergleich benutzt werden kann. In Abbildung 4.1 ist ein typisches Beispiel dargestellt. Hier wird die Qualität einer Encodierung bewertet. Das decodierte Video kann mit dem Eingangssignal des Encoders verglichen werden. Es wird deutlich, dass nur ein Teil der gesamten Kette evaluiert wird.

Vergleicht man dagegen unterschiedlich generierte 3D-Videoobjekte auf der Ebene der generierten Ansichten, wird eine Evaluierung der gesamten Kette vorgenommen (siehe Abbildung 4.2). An diesem Beispiel lässt sich die Beantwortung der Referenzfrage erläutern. Man kann als Referenz für ein 3D-Videoobjekt ein anderes 3D-Videoobjekt (desselben Aufnahmeobjektes) benutzen, welches keine Störungen bzw. eine ausreichende Qualität aufweist (siehe Abbildung 4.3). Auch wenn es nicht immer offensichtlich ist, wird diese Forderung auch bei der 2D-Videoqualitätsbeurteilung angewendet (z. B. wie in Abbildung 4.1). Diese Referenz ist ebenso nicht *ideal*, da natürlich auch diese Aufnahme in ihren Qualitätsparametern (z. B. Auflösung, Rauschen) begrenzt ist. Sie wird aber als *ausreichend gut* angesehen, um so als Referenz zu dienen und um die relevanten Störungen des Testsignals erfassen zu können.

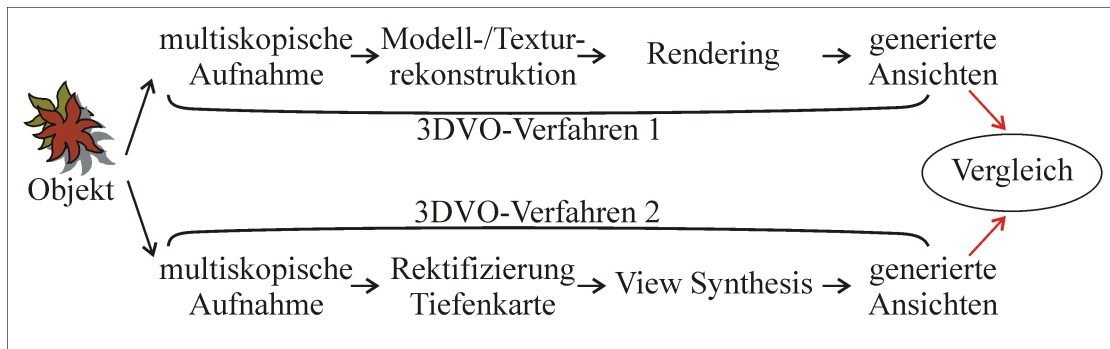


Abbildung 4.2: Vergleich von 3D-Videoobjekten

- *Conclusio 4.4*: Als Referenz für ein 3D-Videoobjekt kann ein anderes 3D-Videoobjekt genutzt werden, welches die erforderliche Qualität aufweist.

Selbstverständlich muss diese Referenz sich auf den gleichen Anwendungsfall (Navigation) beziehen (siehe Abschnitt 4.1).

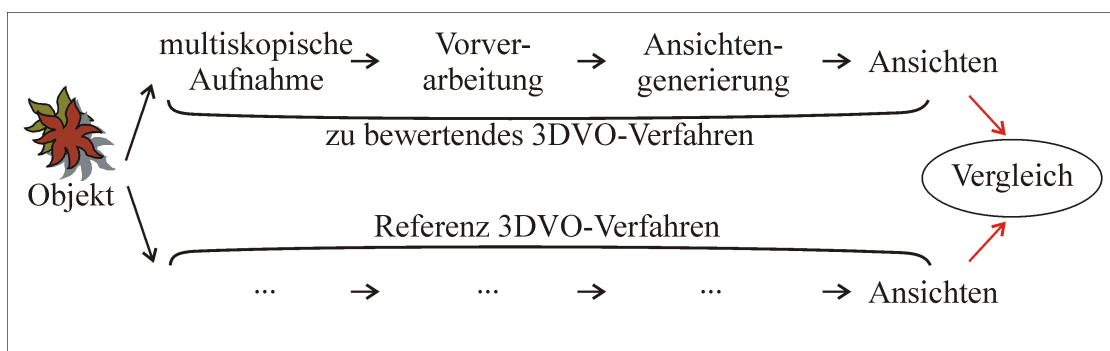


Abbildung 4.3: Referenz für 3D-Videoobjekte

4.3.2 Verwendung zusätzlicher Ansichten

Eine Ergänzung (aber kein Ersatz) für das Referenz-3DVO sind zusätzliche Ansichten des Aufnahmeobjektes, die nicht für die Generierung des 3D-Videoobjektes genutzt wurden. Diese Möglichkeit wird u. a. zum Vergleich verschiedener View-Synthesis-Verfahren genutzt [ISO03b]. Mit Hilfe dieser zusätzlichen Ansichten lassen sich Qualitätsmerkmale, die sich auf genau einen Betrachterstandpunkt beziehen, erfassen. Qualitätsmerkmale, welche sich auf dynamische Eigenschaften während der Navigation beziehen, lassen sich damit nicht ermitteln.

Diese Ergänzung ist mit erhöhtem Aufwand verbunden und natürlich nur dann sinnvoll, wenn ein 3DVO-Verfahren im Entwicklungsstadium evaluiert werden soll. Die zusätzlichen Kameraansichten sind ebenso für die Entwicklung der Qualitätsmerkmalsmodelle hilfreich.

4.3.3 Grenzen der Methode

Mit der oben erläuterten Lösung des Referenzproblems lassen sich nicht alle Aspekte der Qualitätsbewertung von 3D-Videoobjekten untersuchen. Die offensichtlichste Einschränkung ist die begrenzte Verfügbarkeit eines Referenz-3DVO. Natürlich ist der Aufwand für die Generierung von mehreren 3D-Videoobjekten mit dem Ziel der Qualitätsbewertung nur dann gerechtfertigt, wenn sich die Qualität nicht auf eine andere Weise abschätzen lässt. Das ist in erster Linie bei der Entwicklung von 3DVO-Verfahren und bei der Entwicklung der Qualitätsmerkmalsmodelle der Fall.

- *Conclusio 4.5:* Die Verwendung eines Referenz-3DVO ist insbesondere im Entwicklungsstadium von 3DVO-Verfahren sinnvoll.

Die an ein Referenz-3DVO gestellte Forderung einer ausreichenden Qualität (siehe Unterabschnitt 4.3.1) ist nur in einem begrenzten Maße erfüllbar. Grundsätzlich lässt sich die Frage nach einer *ausreichenden* Qualität nur durch den Nutzer, d. h. nur durch subjektive Tests beantworten. Die Beurteilung solcher technischen Darstellungen durch den Menschen ist von dessen Erwartungen und Erfahrungen abhängig. So wird ein Mensch ein solch neuartiges 3D-Videoobjekt auch an Merkmalen des ihm bekannten 2D-Video messen. Auch wird die Bewertung durch den Menschen strenger werden, wenn ihm 3D-Videoobjekte als Medienobjekttyp aus Anwendungen bekannt geworden sind.

- *Conclusio 4.6:* Die Anforderungen an die Qualität des Referenz-3DVO sind durch die Gruppe der Anwender zu bestimmen und sind von Entwicklungen der Technologie, des Nutzungsverhaltens etc. abhängig.

Die vom Referenz-3DVO erfüllten Anforderungen legen den Maßstab für die 3DVO-Qualitätsbewertung fest, was allerdings keine Besonderheit dieser Qualitätsbewertung ist. Hier wird jedoch besonders deutlich, dass die Detektion von Qualitätsmerkmalen von den Eigenschaften der Referenz abhängt. Allerdings sind die Qualitätsmerkmale für 3D-Videoobjekte noch nicht etabliert.

- *Conclusio 4.7:* Mit dem Referenz-3DVO sind die erfassbaren Qualitätsparameter festgelegt.

4.4 Qualitätsmerkmale von 3D-Videoobjekten

4.4.1 Perspektivenfehler

3D-Videoobjekte können von einer abweichenden Perspektive aufgenommen oder dargestellt worden sein. Dieses Merkmal kann bei konventionellem Video nicht auftreten, wohingegen einige Auswirkungen vergleichbar sind. Der Perspektivenfehler des 3D-Videoobjektes kann Folge eines falschen Betrachterstandpunktes/-richtung oder eines falschen Objektortes/-ausrichtung sein. Des Weiteren können auch optische Parameter (Brennweite) abweichen. Perspektivenfehler können folgende Auswirkungen haben:

- Die generierte Ansicht ist verschoben (\rightarrow Translationsfehler).
- Das Videoobjekt weist eine Skalierung auf, die bei falscher Brennweite tiefenabhängig ist.
- Das Videoobjekt ist unter einem falschen Blickwinkel zu sehen (Abbildung 4.4 ⁴).



Abbildung 4.4: Referenz-3DVO (links) und 3DVO mit Perspektivenfehler (rechts)

4.4.2 Silhouettenfehler

3D-Videoobjekte weisen im Gegensatz zu konventionellem Video in jeder generierten Ansicht eine beliebige Form, d. h. eine Silhouette auf (siehe Abschnitt 3.1). Auch diese Silhouette kann Störungen an der Kontur oder im Inneren aufweisen. In Abbildung 4.5 ist ein Beispiel für Störungen entlang der Kontur einer Silhouette abgebildet⁵.

Störungen innerhalb der Silhouette beziehen sich auf Bereiche im Inneren, welche nicht zum Objekt gehören, d. h. durchsichtig sind⁶. Die Fehler des Silhouetteninnern können derart sein, dass die inneren Öffnungen fehlen, zusätzliche Öffnungen existieren⁷ oder die Öffnungen in ihrer Form gestört sind.

⁴Testobjekt Nr. 3, R. Zimmermann, TU Ilmenau

⁵MPEG-Testset „Lara“, Rendering: ETH Zürich (Schweiz)

⁶Das ist zum Beispiel der Fall, wenn das 3D-Videoobjekt eine Person darstellt, welche die Hand in der Hüfte hält und der Bereich zwischen Arm und Oberkörper dadurch durchscheinend ist. (siehe ungestörte Silhouette in Abbildung 4.12)

⁷siehe gestörte Silhouette in Abbildung 4.12



Abbildung 4.5: Silhouette eines Referenz-3DVO (links) und eine gestörte Silhouette (rechts)

Eine weitere Klasse an Silhouettenfehlern ist eine fehlerhafte Transparenz, insbesondere im Randbereich. Im Randbereich weisen Silhouetten oft eine Teiltransparenz auf. Diese erzielt einen visuellen Eindruck, welcher durch die optischen Phänomene an den Objektkanten (Beugung und Spiegelung) bedingt ist. Ein typischer Fehler ist ein zu großer oder ein zu kleiner Randbereich.

4.4.3 Synthesespezifische Fehler

Naturgemäß entstehen bei jedem 3DVO-Verfahren Fehler, die spezifisch für das jeweilige Verfahren sind. So gibt es beispielsweise Warping- oder texturierungsspezifische Fehler. Beim Warping werden Bildpunkte entlang der Epipolarlinien verschoben. Hierfür muss der korrespondierende Bildpunkt der zweiten Ansicht bekannt sein (siehe Unterabschnitt 3.2.3). Wird die Korrespondenz falsch ermittelt, wird der Punkt auf der Epipolarlinie an eine falsche Position verschoben und es kommt zu den typischen Verzerrungen entlang der Epipolarlinien (siehe Abbildung 4.6 ⁸).



Abbildung 4.6: Referenz-3DVO (links) und 3DVO mit Warping-spezifischen Fehlern (rechts)

Bei den modellbasierten 3DVO-Verfahren können u. a. Fehler durch die Texturierung und die spezifische Kompression der Texturdaten auftreten. Diese äußern sich z. B. durch Objektbereiche, welche auffällig undeutlich oder verzerrt wirken. In Abbildung 4.7 ist

⁸Objekt „IAVAS“, M. Duelli, TU Ilmenau [Due03]

ein Referenz-3DVO und ein 3D-Videoobjekt⁹ mit Fehlern in der Textur abgebildet. Diese rühren von einer zu hohen Kompression der Texturdaten (*Average Texture Compression*).



Abbildung 4.7: Referenz-3DVO (links) und 3DVO mit texturierungsspezifischen Fehlern (rechts)

Natürlich gibt es viele weitere synthesespezifische Fehler, die abhängig vom jeweils verwendeten Verfahren sind. Da mit Weiterentwicklungen und auch mit neuen Verfahren zu rechnen ist, sind auch neue Arten von Fehlern zu erwarten. Gleichzeitig muss man aber davon ausgehen, dass mit den Weiterentwicklungen die Fehler weniger apparent werden.

4.4.4 Videospezifische Fehler

Für die Generierung von 3D-Videoobjekten verwenden viele Verfahren zweidimensionales Video als Ausgangsmaterial. Dies gilt sowohl für die bildbasierten Verfahren, aber auch modellbasierte Verfahren nutzen Video als Textur. Daher treten Qualitätsmerkmale von konventionellem Video auch bei diesen 3D-Videoobjekten auf. Störungen wie Unschärfe und Blockartefakte beeinträchtigen in ähnlicher Weise die Qualität von 3D-Videoobjekten.

4.4.5 Verdeckungsfehler (Okklusionen)

Wenn Bereiche des Aufnahmeobjektes während der Aufnahme verdeckt waren, sind sie im Videoobjekt nicht korrekt darstellbar und es kommt so zu den so genannten Okklusionen. Die 3DVO-Verfahren versuchen auf unterschiedliche Weise, die Okklusionen zu retuschieren, z. B. indem sie den Bereich mit einer Textur aus der Umgebung füllen. Dennoch mindern sie die Qualität des Videoobjektes.

4.4.6 Dynamik der Qualitätsmerkmale

Die dynamischen Eigenschaften der Qualitätsmerkmale erfordern eine besondere Betrachtung. Im Gegensatz zu konventionellem Video bedarf es einer Unterscheidung der dynamischen Eigenschaften in eine Klasse, die zeitbezogen ist und in eine Klasse, die auf Blickpunktänderungen bezogen ist:

⁹Testobjekt „MPI-Dancer“, Max-Planck-Institut für Informatik, Saarbrücken

1. *zeitabhängig*: Das Qualitätsmerkmal (z. B. Warping-Artefakte, Flackern) ändert sich während der Laufzeit. Diese Eigenschaft weisen Qualitätsmerkmale von konventionellem Video ebenso auf.
2. *blickpunktsänderungsabhängig*: Das Qualitätsmerkmal (z. B. Perspektivenfehler) ändert sich während der Navigation auf oder um das Objekt. Diese Eigenschaft weisen Qualitätsmerkmale von konventionellem Video nicht auf. Die Qualität wird durch diese Erscheinung oft deutlich gemindert, z. B. wenn das Videoobjekt während der Navigation „springt“.

4.5 Ansatz zur Entwicklung von Merkmalsmodellen

4.5.1 Methodik zur objektiven Bewertung der Qualitätsmerkmale

Im vorhergehenden Abschnitt wurden die Qualitätsaspekte von 3D-Videoobjekten beschrieben. Diese Aufstellung beruht auf verschiedenen Untersuchungen und der Offen-sichtlichkeit einzelner Merkmale. Daher sind folgende Aspekte zu berücksichtigen:

- *Vollständigkeit*: Es liegt kein Beweis (z. B. mittels statistischer Untersuchungen) vor, dass die Aufstellung vollständig ist. Des Weiteren können neue 3DVO-Verfahren neue Qualitätsmerkmale aufweisen.
- *subjektive Bewertung*: Die Aufstellung erlaubt keine Aussage, wie die genannten Qualitätsaspekte die subjektive Bewertung beeinflussen.
- *gegenseitige Abhängigkeit*: Aus der Aufstellung lassen sich keine Abhängigkeiten zwischen den Qualitätsmerkmalen bei der Beeinflussung der subjektiven Bewertung ableiten.

Im Abschnitt 4.1 wurde herausgearbeitet, dass eine Regression der Qualitätsmerkmale vorzunehmen ist. Hierfür sind mathematische Modelle erforderlich, mit denen die Merkmale analytisch erfassbar werden.

Eine ähnlich gelagerte Problemstellung wurde in [Kay96] untersucht. Dort wurde ein Modell (Faktorenmodell) der Bildqualität datenkomprimierter stereoskopischer Fernsehbilder nach folgender Abfolge entwickelt:

1. Ermittlung der subjektiven Beurteilungsaspekte mittels subjektiver Tests (Fragebögen mit Vorgaben) auf Basis einer repräsentativen Bilddatenbank
2. Ermittlung der Dimensionen der subjektiven Wahrnehmung mittels Faktoren- und Clusteranalyse
3. Bildung (eindeutig rückführbarer) mathematischer Modelle für die relevanten Aspekte
4. Bildung eines Qualitätsmaßes aus den Modellen mittels der Ergebnisse aus der Faktorenanalyse

Mit dieser Methode konnten die zunächst untersuchten neun Qualitätsaspekte auf drei zu modellierende Aspekte (Schärfe, Rauschen, Blockausfälle) reduziert werden. Für diese drei Aspekte wurden mittels geeigneter Methoden (Frequenzanalyse, Varianzanalyse,

Blockuntersuchung) Modelle zur Messung entwickelt. Die Zusammenführung zu einem Qualitätsmaß erfolgte mit Hilfe der Gewichtungsfaktoren, welche mit der Faktorenanalyse ermittelt wurden.

Für die objektive Bewertung von konventionellem Video wurde in [Wol02] (siehe auch Unterabschnitt 2.4.2) folgende Abfolge beschrieben:

1. Ermittlung von Merkmalen (ähnlich ANSI T1.801) auf Basis statistischer Größen
2. Bildung von Qualitätsparametern aus den Merkmalen mit Hilfe verschiedener Funktionen (Vergleichsfunktionen, Skalierungen, Abstandsmaße, statistische Maße, Beschneidungsfunktionen).
3. gewichtete Summation (in Anlehnung an subjektive Tests) von Qualitätsparametern zu anwendungsspezifischen Videoqualitätsmaßen

In [Wol02] wird zunächst kein unmittelbarer Zusammenhang zwischen den subjektiven Beurteilungsaspekten und den Qualitätsparametern dargelegt¹⁰. Es ist davon auszugehen, dass es bei den statistischen Maßen einige Überschneidungen zwischen den Qualitätsparametern und den Beurteilungsaspekten gibt. Solche Überschneidungen werden in [Kay96] vermieden. Das Vorgehen in [Kay96] ist möglich, da dort auf Grund des eingeschränkten Fehlerbildes nur wenige Beurteilungsaspekte zurückgeführt werden müssen. Ist dies nicht der Fall, gilt für eine Vielzahl der genutzten (statistischen) Methoden: Ein Parameter p_i ist von mehreren Qualitätsaspekten/-merkmalen $\Psi_1, \Psi_2 \dots \Psi_N$ abhängig. Dies liegt darin begründet, dass die Merkmalsgruppe $\Psi_1, \Psi_2 \dots \Psi_N$ ähnliche (statistische) Auswirkungen hat.

$$p_i = f_x(\Psi_1, \Psi_2 \dots \Psi_N) \quad (4.1)$$

Umgekehrt kann sich ein Qualitätsmerkmal Ψ_i auf mehrere Parameter $p_0, p_1, p_2 \dots p_N$ auswirken.

$$p_0 = f_0(\Psi_i), p_1 = f_1(\Psi_i), p_2 = f_2(\Psi_i) \dots, p_N = f_N(\Psi_i) \quad (4.2)$$

Die Ursache hierfür ist, dass mittels unterschiedlicher Methoden die Spezifika eines Merkmals Ψ_i erfasst werden.

Für die objektive Bewertung der Qualitätsmerkmale und -aspekte von 3D-Videoobjekten sind folgende Voraussetzungen gegeben:

- Es gibt im Vergleich zur Bewertung von konventionellem Video sehr viele Qualitätsmerkmale und -aspekte. Dies erschwert die Rückführung der Parameter auf die Merkmale.
- Die Bestimmung der Dimensionen der subjektiven Wahrnehmung (siehe [Kay96]) erfordert eine bekannte und abgegrenzte Menge an Fehlern und Störungen. Diese Voraussetzung ist für 3D-Videoobjekte auf Grund des Entwicklungsstandes nur teilweise gegeben.
- Für die Bewertung von konventionellem Video kommt ein breites Spektrum an Methoden zum Einsatz (siehe Abschnitt 2.4). Ein wichtiger Unterschied bei der Bewertung von 3D-Videoobjekten sind perspektivische Fehler, welche einen bildpunktweisen Vergleich mit der Referenz verhindern. Es ist daher eine in grundlegenden Aspekten neue Methodik erforderlich.

¹⁰Mit einer *Impairment RCA* (root cause analysis) werden Methoden zur Zuordnung von Parametern zu einzelnen Störungstypen gesucht.

Im Ergebnis der Analyse der speziellen Gegebenheiten für die objektive Bewertung der Qualitätsmerkmale und -aspekte von 3D-Videoobjekten wird folgender Ansatz gewählt:

1. Entwicklung von ein oder mehreren Maßen w_i, w_j, \dots für die festgestellten Qualitätsaspekte und -merkmale Ψ_k
2. Zuordnung von Merkmalsgruppen $\Psi_1, \Psi_2 \dots \Psi_N$
3. ggf. ein Vergleich der Maße w_i, w_j, \dots eines Merkmals Ψ_k
4. ggf. eine Untersuchung eines Maßes w_i in Bezug auf verschiedene Merkmale $\Psi_1, \Psi_2 \dots \Psi_N$

Es ist selbstverständlich, dass die Entwicklung von Merkmalsmaßen w_i ein Prozess mit mehreren Iterationen ist. Daher sind die genannten Schritte 1...4 nach der Entwicklung und Überprüfung eines 3D-Videoobjektqualitätsmaßes zu wiederholen.

4.5.2 Bildung von Merkmalsmaßen

Im Abschnitt 4.6 werden Maße mit Hilfe der vorgestellten Methoden für die verschiedenen Qualitätsmerkmale (siehe Abschnitt 4.4) abgeleitet. Diese Merkmalsmaße sind Grundlage zur Bildung von Qualitätsparametern und letztlich des 3D-Videoobjektqualitätsmaßes (siehe Kapitel 5). Im ersten Schritt werden die Merkmale während eines Zeitpunktes T untersucht, das heißt es werden einzelne Ansichten¹¹ als Untersuchungsgegenstand genutzt. Für die Bildung der Merkmalsmaße wird folgender Ansatz gewählt: Für ein Merkmal Ψ wird mittels einer Methode χ ein Merkmalsmaß $w_{\Psi,\chi}$ bestimmt.

$$w_{\Psi,\chi} = f_{\chi}(\Psi) \quad (4.3)$$

Für den Fall, dass eine Methode χ mehrere Merkmale $\Psi_1, \Psi_2 \dots \Psi_N$ erfasst, gilt:

$$w_{\chi} = f_{\chi}(\Psi_1, \Psi_2 \dots \Psi_N) \quad (4.4)$$

Folgende Anforderungen werden an die Merkmalsmaße $w_{\Psi,\chi}$ gestellt:

- Das Merkmalsmaß $w_{\Psi,\chi}$ steht im stetigen Zusammenhang zum Merkmal Ψ oder zur Merkmalsgruppe $\Psi_1, \Psi_2 \dots \Psi_N$
- Ist das Merkmal Ψ (d. h. die Störung) oder die Merkmalsgruppe $\Psi_1, \Psi_2 \dots \Psi_N$ klein bzw. nicht vorhanden, strebt $w_{\Psi,\chi} \rightarrow 0$
- Die Merkmalsmaße $w_{\Psi,\chi}$ erfassen noch nicht den Einfluss auf die subjektive Wahrnehmung. Diese Anpassung erfolgt im Schritt der Bildung der Qualitätsparameter p und letztlich bei der Bildung des 3D-Videoobjektqualitätsmaßes.

Diese Anforderungen zeigen, dass die Merkmalsmaße $w_{\Psi,\chi}$ noch nicht die subjektiven Bewertung erfassen¹². Dennoch wird sich diese Trennung nicht vollständig umsetzen lassen, da die Maße an sich die (vom Menschen bestimmten) Qualitätsmerkmale beschreiben sollen und bereits bei der Entwicklung der Merkmalsmodelle Wahrnehmungsaspekte beachtet werden müssen.

¹¹im Sinne von Einzelbild, vergleichbar mit dem Begriff *video object plane* für beliebig berandete Videoobjekte

¹²Diese Anforderung steht an die Qualitätsparameter p .

4.5.3 Untersuchung der Dynamik von Merkmalsmaßen

Die Merkmalsmaße $w_{\Psi, \chi}$ bzw. w_{χ} müssen zur Bildung von Qualitätsparametern des Videoobjektes einschließlich ihrer dynamischen Eigenschaften betrachtet werden. Im Unterabschnitt 4.4.6 wurden die beiden Klassen der dynamischen Eigenschaften von Qualitätsmerkmalen eingeführt. Zudem können Eigenschaften beider Klassen gemeinsam auftreten¹³, was in der Modellbildung zu berücksichtigen ist. Für die Unterscheidung der dynamischen Eigenschaften ergibt sich folgende Problemstellung: Da sich die Störungen der beiden Klassen nicht grundlegend unterscheiden, ist eine Zuordnung mittels objektiver Verfahren nicht trivial. Jedoch ist eine solche Zuordnung durch den menschlichen Betrachter auf Grund von Erfahrungswerten in vielen Fällen möglich.

- *Beispiel 1:* Während einer Navigation um ein Objekt herum kommt es häufig zu Störungen, wenn sich die Betrachterposition zwischen den Positionen der ursprünglichen Aufnahmekameras befindet. Das Videoobjekt zeigt dann beispielsweise Texturierungsfehler oder Silhouettenverzerrungen. Für den Betrachter mit Erfahrungswerten ist dieses Verhalten während der Navigation bekannt. Er wird annehmen, dass es Störungen auf Grund der Navigation sind. Um eine solche Einschätzung mit Hilfe eines objektiven Verfahrens zu treffen, muss der Navigationspfad bekannt sein und das Verhalten der Störung diesem Pfad zuordenbar sein.
- *Beispiel 2:* Wenn sich das Aufnahmeobjekt in größerem Maße bewegt oder sich ändert, können Störungen auftreten (z. B. auf Grund fehlerhaft erkannter Korrespondenzen). Diese Störung gehört zu der Klasse der zeitabhängigen Störungen. Wenn man in diesem Falle um das Objekt herum navigiert, treten dann oft zusätzlich blickpunktsabhängige Störungen auf, weil die Störanfälligkeit auf den Positionen zwischen den Aufnahmekameras steigt.

Diese Beispiele zeigen, dass eine objektive Unterscheidung in Zeitabhängigkeit und Blickpunktsänderungsabhängigkeit auf Basis der Störungen viele Probleme aufwirft. Daher steht die Frage, für welche Aspekte diese Unterscheidung notwendig ist:

- Da die Auswirkungen der beiden Eigenschaftsklassen auf die Störungen des Videoobjektes ähnlich sind, können sie mit denselben Methoden erfasst werden. Die Resultate dieser Messungen sind die verschiedenen dynamischen Merkmalsmaße.
- Für die Bildung von Qualitätsparametern ist ein Bezug zum Navigationspfad erforderlich. Man kann ihn jedoch für die Qualitätsbeurteilung als bekannt voraussetzen, da er grundsätzlich für eine Nutzung des 3DVO bekannt sein muss.

Der Bezug eines Merkmalsmaßes auf den Navigationspfad erlaubt die Berücksichtigung der veränderten visuellen Wahrnehmung.

- *Conclusio 4.8:* Eine Unterscheidung der dynamischen Qualitätsmerkmale erfolgt nach der Bildung der dynamischen Merkmalsmaße durch einen Bezug auf den gegebenen Navigationspfad.

¹³Eine Blickpunktsänderung setzt auch eine Zeitänderung voraus. Daher treten die dynamischen Eigenschaften einer Blickpunktsänderung nie separat auf.

4.5.4 Auswahl geeigneter Methoden

Korrespondenzanalyse

Da man bei der 3DVO-Qualitätsbewertung davon ausgehen muss, dass das Videoobjekt einem perspektivischen Fehler unterliegt, benötigt man Informationen über die örtliche Zuordnung zur Referenz. Daher ist eine Korrespondenzanalyse notwendig. Hierfür gibt es u. a. folgende Methoden: Merkmalsextraktion, Block-Matching, Abschätzung des optischen Flusses [Red00].

Für eine Merkmalsextraktion müssen die für die Detektion geeigneten Merkmale definiert sein. Hierfür werden meist Kanten benutzt. Kanten, wie die Silhouette eines 3D-Videoobjektes, lassen sich leicht finden¹⁴, während sich im Inneren des Videobjektes nicht immer Kanten finden lassen.

Beim Block-Matching wird das Bild in Blöcke unterteilt und ein Vektor für jeden Block gesucht, der zu einem möglichst ähnlichen Block führt. Das Kriterium für die maximale Ähnlichkeit ist das Minimum der SAD oder der SSD¹⁵. Bei 3D-Videoobjekten variiert die interessierende Blockgröße: Bei einem Translationsfehler ist das gesamte Bild als Block zu untersuchen. Bei Warping-Fehlern sind die Blöcke wenige Bildpunkte groß. Daher bietet sich das hierarchische Block-Matching an. Ein Translationsfehler würde sich dann in der Ebene 0 und ein Warping-Fehler in einer hohen Ebene erkennen lassen. Weitere Möglichkeiten ergeben sich durch die allgemeine Form des Block-Matching, bei der zu einem gegebenen (rechteckigen) Block ein Viereck (d. h. dessen Ecken) gesucht wird, welches maximal ähnlich ist¹⁶. Hierdurch können sich Perspektivenfehler erkennen lassen.

Methoden, welche auf der Abschätzung des optischen Flusses basieren, erfordern eine (zeitlich und örtlich) kontinuierliche Änderung des Grauwertes. Da dies bei Störungen von 3D-Videoobjekten nicht immer der Fall ist, lassen sich solche Verfahren nur für einige Sonderfälle einsetzen.

Konturtransformation

Die Silhouette ist ein wichtiges Merkmal eines 3D-Videoobjektes und wird durch die Alpha-Komponente beschrieben. Um eine Silhouette mit deren Referenz zu vergleichen, ist ein bildpunktweiser Vergleich der Alpha-Werte wegen der Perspektivenfehler nicht möglich. Als Alternative bietet sich die Verwendung der Kontur der Silhouette an. Hierfür gibt es mehrere Möglichkeiten zur Repräsentation. Für die Fourier-Approximation wird die Kontur zunächst als eindimensionale komplexe Funktion beschrieben:

$$t(p) = n(p) + j \cdot m(p) \quad \text{mit: } 0 \leq p \leq P \quad (P: \text{Anzahl der Punkte}) \quad (4.5)$$

Die komplexen Koeffizienten der (eindimensionalen) Diskreten Fourier-Transformation ergeben sich zu:

$$T(q) = \sum_{p=0}^{P-1} t(p) \cdot e^{-j \cdot \frac{2 \cdot \pi \cdot p \cdot q}{P}} \quad (4.6)$$

Diese Koeffizienten weisen Eigenschaften auf, die für den Vergleich einer Kontur mit ihrer Referenz genutzt werden können. Eine Translation der Kontur bewirkt ausschließlich

¹⁴Die Silhouette ist bereits durch die Maskeninformation (Alpha-Werte) bekannt.

¹⁵sum of absolute differences, sum of squared differences

¹⁶vergleichbar mit dem Texturierungsprozess auf einem 2D-Gitter

eine Änderung von $T(0)$. Eine Skalierung der Kontur bewirkt eine gleichmäßige Skalierung aller Koeffizienten $T(q)$ mit $q > 0$. Eine Rotation der Kontur äußert sich in einer Phasendrehung der Koeffizienten $T(q)$ mit $q > 0$ um denselben Betrag [Ohm99b].

Eine weitere Möglichkeit zum Vergleich von Konturen ergibt sich durch die Nutzung der Wavelet-Approximation. Durch die Tiefpassfilterung halbiert sich jeweils die Anzahl der Koeffizienten, wobei die letzte Transformationsebene $N = 4 \dots 16$ Werte enthalten sollte [Ohm99b]. Die Koeffizienten der Wavelet-Transformation weisen ähnliche Eigenschaften wie die der DFT auf. Eine Translation bewirkt eine unmittelbare Addition der x - und y -Komponente des Verschiebungsvektors jeweils zu den x - und y -Koeffizienten der letzten Transformationsebene. Eine Skalierung der Kontur skaliert gleichmäßig alle Wavelet-Koeffizienten. Ein Unterschied zur DFT ist der lokale Bezug der Koeffizienten. Daher werden beispielsweise durch eine kleine lokale Änderung an der Kontur nur die zugeordneten hochfrequenten Koeffizienten geändert. Bei der Verwendung von kartesischen Koordinaten sind beide Koordinaten gleichbedeutend für die Erkennung der Kontur. Für Polarkoordinaten ist die Betragskoordinate weit wichtiger, sodass die Phasenkoordinate vernachlässigt werden kann [Ohm99b], [Mül99].

Bei der Konturbildung muss jedoch eine Eigenschaft der 3D-Videoobjektansichten vernachlässigt werden. Die Silhouette kann teiltransparent sein, da die Alpha-Information den gleichen Wertebereich wie die anderen Kanäle aufweist (*grey scaled shape*). Bei der Konturbildung muss daher eine Binarisierung vorgenommen werden. Eine Möglichkeit ist die Kontur als äußerste Begrenzung der Silhouette aufzufassen. Sinnvoller erscheint jedoch die Verwendung eines Schwellenwertes¹⁷.

Methoden der Segmentierungsqualitätsbewertung

Die Silhouette eines 3D-Videoobjektes gleicht in mehreren Aspekten der Maske eines segmentierten zweidimensionalen Videoobjektes. Also liegt es nahe, Methoden zur relativen Bewertung der Segmentierungsqualität (siehe 2.4.4) auch zur Bewertung von Silhouetten von 3D-Videoobjekten einzusetzen. Da diese Methoden aber eine bildpunkt-basierte Repräsentation nutzen, dürfen Perspektivfehler nicht vorhanden sein oder müssen bereits korrigiert worden sein.

Örtlicher und zeitlicher Informationsgehalt

Der örtliche und der zeitliche Informationsgehalt SI , TI bilden die Basis für statistische Parameter zur Bewertung von Qualitätsmerkmalen wie z. B. Verschmierungen von konventionellem Video. Einige Merkmale von 3D-Videoobjekten verändern diese Parameter in ähnlicher Weise: Verzerrungen entlang der Epipolarlinien, Bereiche mit fehlerhaft vergebener Textur etc.

Histogrammvergleich

Farbhistogramme repräsentieren die Häufigkeitsverteilung der Farbwerte eines Bildes oder einer Bildsequenz. Sie werden also von der Wahl des verwendeten Farbraumes (RGB, HSV etc.) bestimmt. In einem ersten Ansatz ergibt sich ein einkanaliges Farbhistogramm H für eine RGB-Repräsentation aller Bildpunkte M (Zeile i , Spalte j) nach folgender Vorschrift:

$$M(i, j) = \begin{pmatrix} r \\ g \\ b \end{pmatrix} \Rightarrow \text{inc}(H(r, g, b)) \quad (4.7)$$

¹⁷Aus den Erfahrungen bei der Codierung von beliebig geformten Videoobjekten ergab sich 16 als eine geeignete Schwelle (bei 8 Bit Quantisierung).

Dieser Ansatz ist der genaueste und gleichzeitig der aufwändigste. Um den Aufwand zu mindern, verwendet man in der Bilderkennung einen Satz an festzulegenden Referenzfarben (z. B. 64, 112 oder 256 Referenzfarben). Für die Histogrammbestimmung muss für die Farbwerte aller Punkte $M(i, j)$ entschieden werden, welcher Referenzfarbe $F_k = (r_k, g_k, b_k)^T$ sie zuzuordnen sind. Dies kann beispielsweise mit dem minimalen euklidischen Distanzmaß $d = \sqrt{(r - r_k)^2 + (g - g_k)^2 + (b - b_k)^2}$ erfolgen. Alternativ zur Referenzfarbzuzuordnung kann eine Histogramminimierung durch eine gröbere Quantisierung der Farbkanäle erfolgen.

In der Bilderkennung geht man von der Annahme aus, dass die Ähnlichkeit von Histogrammen eine Folge der Ähnlichkeit der untersuchten Bilder ist. Zur Feststellung dieser Ähnlichkeit werden u. a. verschiedene Maße der Clusteranalyse verwendet. Dies sind die Maße der Minkowski-Metrik (City-Block-Distanz d_{L1} , euklidische Distanz d_{L2} ; siehe Gleichungen 4.8) oder skaleninvariante Maße wie die Mahalanobis-Metrik.

$$d_{L1}(H_{ref}, H_{eval}) = \sum_{i=1}^N |H_{ref}(i) - H_{eval}(i)|$$

$$d_{L2}(H_{ref}, H_{eval}) = \sqrt{\sum_{i=1}^N (H_{ref}(i) - H_{eval}(i))^2} \quad (4.8)$$

Eine weitere Methode ist der Histogrammvergleich (*histogram intersection*). Der Index V dient dabei als Hauptvergleichsmerkmal und berechnet sich für die Farbhistogramme H_{eval} und H_{ref} (des zu bewertenden Objektes und der Referenz) wie folgt¹⁸:

$$V(H_{ref}, H_{eval}) = \frac{\sum_{b=0}^{255} \sum_{g=0}^{255} \sum_{r=0}^{255} \min(H_{ref}(b, g, r), H_{eval}(b, g, r))}{\sum_{b=0}^{255} \sum_{g=0}^{255} \sum_{r=0}^{255} H_{eval}(b, g, r)} \quad (4.9)$$

Der Index V ist kaum variant bei Änderungen der Auflösung oder der Orientierung, aber stark abhängig von Beleuchtungsänderungen [Haf98].

4.6 Entwicklung von Merkmalsmodellen

In diesem Abschnitt werden Modelle für verschiedene Qualitätsmerkmale vorgestellt. Die resultierenden Maße $w_{\Psi, \chi}$ bzw. w_{χ} beziehen sich auf jeweils eine einzelne Ansicht.

4.6.1 Translations- und Skalierungsfehler

Für Methoden zur Bestimmung von Translations- und Skalierungsfehlern gelten einige Aspekte, wie sie auch für die Bilderkennung oder die Bewegungsabschätzung (für die Videodatenkompression) gelten. Daher kommen Methoden wie das Block-Matching und die Kontur-Transformation zum Einsatz. Für eine schrittweise Problemlösung werden zunächst die separat auftretenden Fehler untersucht.

¹⁸für einen RGB-Farbraum mit 8 Bit Quantisierungsbreite je Kanal

Translationsfehler

Translationsfehler treten meist auf, wenn die Parameter des Kamera-Setup falsch gemessen (bei fester Kalibrierung) oder falsch berechnet (bei Selbstkalibrierung) worden sind. Der für die Bestimmung dieses Fehlers einfachste Fall liegt vor, wenn ausschließlich ein solcher Translationsfehler vorliegt. In einem solchen Fall kann mit dem hierarchischen Block-Matching auf der Ebene 0 oder der Konturtransformation eine Detektion erfolgen. Für einen simulierten Translationsfehler¹⁹ wurden diese Verfahren getestet. Als Kriterium für das Block-Matching wurde das Minimum der SSD gewählt. Für diesen Trivialfall wurde die Translation sicher für eine beliebige Komponente (Y, Grün, Alpha) erkannt. Für die Bildung eines Merkmalsmaßes $w_{trans, BM-h}$ wird der Betrag der Summe aller Vektoren \vec{v}_i ins Verhältnis zur Anzahl der Vektoren gesetzt. Dieses Verhältnis wird groß, wenn eine hoher Anteil gleich ausgerichteteter Verschiebungsvektoren ermittelt wurde. Die Beträge der Vektoren entsprechen der Verschiebungsweite in Bildpunkten. Um das Merkmalsmaß auflösungsunabhängig anzugeben, ist es auf ein Referenzmaß d_{ref} zu beziehen. Im allgemeinen Fall ist dies der maximale Durchmesser des Objektes. Es kann aber auch die Höhe bzw. Breite der Projektionsfläche angenommen werden, da sich diese an der maximalen Ausdehnung des Objektes orientiert. Das Merkmalsmaß $w_{trans, h}$ innerhalb einer Ansicht (in der Hierarchieebene h) ergibt sich demzufolge zu:

$$w_{trans, BM-h} = \frac{1}{N \cdot d_{ref}} \left| \sum_{i=0}^N \vec{v}_i \right| \quad (4.10)$$

mit h : Hierarchieebene, N : Anzahl der Objektblöcke, wobei $N \leq 2^h$,
 \vec{v}_i : Verschiebungsvektor, d_{ref} : Referenzmaß,

Reine Translationen werden in der Hierarchieebene 0, d. h. mittels $w_{trans, BM-0}$ quantitativ erfasst.

Eine ausschließliche Translation lässt sich ebenfalls mit Hilfe der Konturtransformation ermitteln. In Abbildung 4.8 sind die Beträge der Koeffizienten $r(0 \dots 10)$, u. a. für die Referenz und die verschobene Kontur dargestellt. Es wird deutlich, dass die Translation ausschließlich eine Änderung von $|T(0)|$ bewirkt. Aber auch beispielsweise eine Skalierung führt zu einer Änderung von $T(0)$, da sich die Anzahl P der Punkte der Kontur unterscheidet. Daher wird $|T(0)|$ in den weiteren Betrachtungen auf die Punktzahl P normiert:

$$|T'(0)| = \frac{|T(0)|}{P} \quad (4.11)$$

Der Parameter $w_{trans, kontur}$ misst diese Translationen als Verhältnis der Differenz der $|T'(0)|$ zu $|T'_{ref}(0)|$.

$$w_{trans, kontur} = \frac{|T'_{eval}(0)| - |T'_{ref}(0)|}{|T'_{ref}(0)|} \quad (4.12)$$

Da im Folgenden ausschließlich die Beträge der komplexen Koeffizienten $T(i)$ betrachtet werden, wird aus Gründen der Übersichtlichkeit auf die Betragsstriche verzichtet:

$$w_{trans, kontur} = \frac{T'_{eval}(0) - T'_{ref}(0)}{T'_{ref}(0)} \quad (4.13)$$

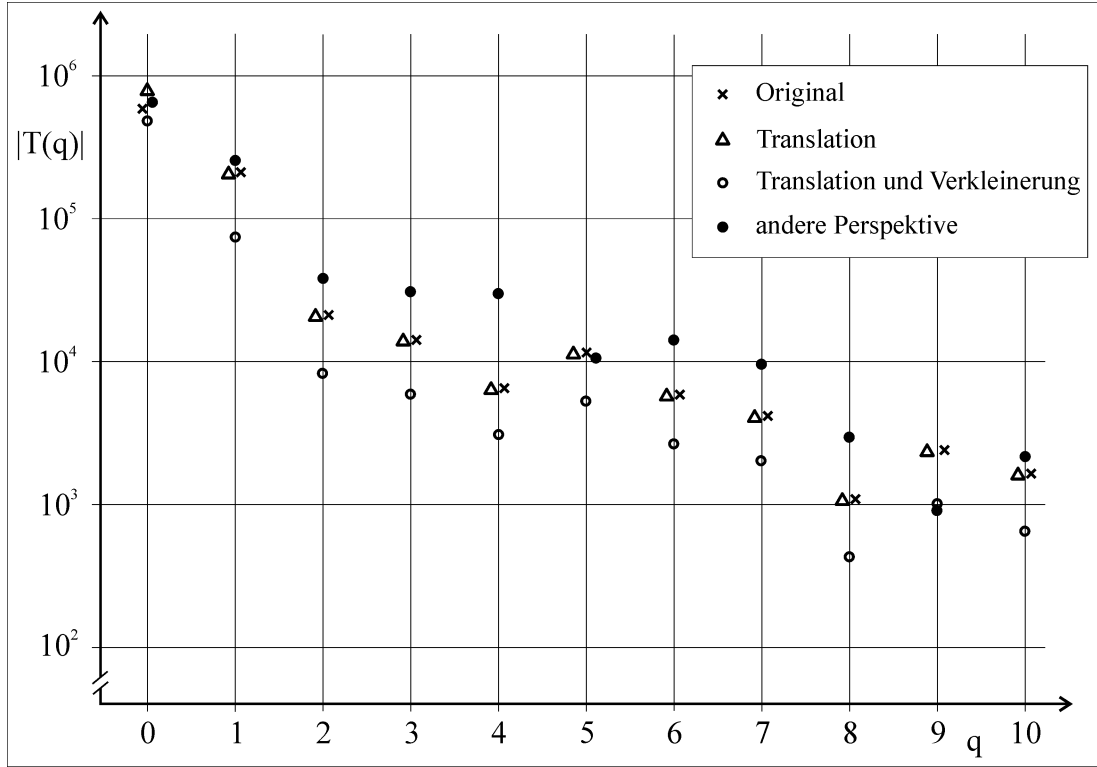


Abbildung 4.8: DFT-Koeffizienten für verschiedene Konturen

Skalierungsfehler

Skalierungsfehler treten meist infolge eines Abstands- oder Brennweitenfehlers auf. Die Ursachen dafür sind ähnlich wie die für Translationsfehler. Ein Abstandsfehler unterscheidet sich von einem Brennweitenfehler darin, dass die Skalierung tiefenabhängig ist.

Die Konturtransformation ermöglicht eine Erkennung eines reinen Skalierungsfehlers. Die Abbildung 4.8 verdeutlicht die Skalierung der DFT-Koeffizienten im Falle einer Skalierung der Kontur²⁰. Im allgemeinen Fall sind weitere Störungen vorhanden, sodass die Skalierung mittels linearer Regression zu ermitteln ist. Hierfür ist die Regression nicht für alle P Koeffizienten notwendig, weil wegen der vernachlässigbar kleinen hochfrequenten Koeffizienten nur eine reduzierte Anzahl P_{red} zu berücksichtigen ist. Für die lineare Regression 4.14 ergeben sich der Regressionskoeffizient b und die Regressionskonstante a nach 4.15 und 4.16.

$$\hat{T}_{eval}(i) = b \cdot T_{ref}(i) + a \quad (4.14)$$

$$b = \frac{\sum_{i=1}^{P_{red}} (T_{ref}(i) - \bar{T}_{ref}) \cdot (T_{eval}(i) - \bar{T}_{eval})}{\sum_{i=1}^{P_{red}} (T_{ref}(i) - \bar{T}_{ref})^2} \quad (4.15)$$

$$a = \bar{T}_{eval} - b \cdot \bar{T}_{ref} \quad (4.16)$$

Die Regressionskonstante a ist für eine geänderte Konturform $\neq 0$, aber für die Detektion

¹⁹Testobjekt „MR“, N. Herzog, TU Ilmenau [Her03]

²⁰Auf Grund der logarithmischen y -Achse wirkt sich der Skalierungsfaktor in einer konstanten Differenz aus.

von Skalierungen nicht zu berücksichtigen. Ein erster Ansatz zur Überprüfung der Regression ist die Bestimmung eines Korrelationskoeffizienten. Der empirischer Korrelationskoeffizient nach Pearson setzt intervallskalierte Variablen voraus. Dies ist für Transformationskoeffizienten nicht gegeben. Ebenso ist die Rangkorrelation für die Überprüfung der Skalierung ungeeignet. Da die Transformationskoeffizienten ungleichmäßig verteilt sind, werden die (quadratischen) Differenzen von $\hat{T}_{eval}(i)$ zu $T_{eval}(i)$ ins Verhältnis zur Größe von $T_{eval}(i)^2$ zu setzen. Das arithmetische Mittel dieser Relationen dient als ein erster Ansatz für ein geeignetes Maß e zur Bestimmung der Abweichungen der ermittelten Regression:

$$e = \frac{1}{P_{red}} \sum_{i=1}^{P_{red}} \frac{\left(\hat{T}_{eval}(i) - T_{eval}(i)\right)^2}{T_{eval}(i)^2} \quad (4.17)$$

Fehler des zu evaluierenden Objektes	Regressions- koeffizient b	Regressions- konstante a	Abweichung e
synthetisierte Skalierung	0,444	-0,54	0,0968
grobe Konturfehler	1,12	378,4	128
synth. Skalierung und Translation	0,444	-0,54	0,0968
Ansichtenfehler	1,18	1295	160
leichte Warping-Fehler	1,00	86,9	4,5
größere Warping-Fehler	1,07	-187	0,96
Texturkompressionsartefakte	1,13	24,7	0,59
Texturierungsfehler	1,00	0,0	0,00

Tabelle 4.1: Lineare Regression von Skalierungsfehlern und deren Bewertung

In Tabelle 4.1 sind für verschiedene Fehlerfälle die Ergebnisse der Regression und das zugehörige Maß e für die Regressionsabweichung aufgelistet. Erwartungsgemäß werden synthetisierte Skalierungsfehler mit einer geringen Regressionsabweichung erkannt. Deutlich wird auch, dass Ansichten- oder Konturfehler zu einem hohem Regressionsabweichungsmaß führen und daher als Skalierung nicht zu berücksichtigen sind. Im Beispiel der Texturkompressionsartefakte führte die hohe Kompression durch ein Ausfransen der Objektkanten tatsächlich zu einer Vergrößerung des Objektes²¹. Im Gegensatz dazu haben sich beim untersuchten Beispiel mit Fehlern auf Grund der Texturcodierung diese nicht auf die Silhouette ausgewirkt²², sodass das Nichtvorhandensein einer Skalierung mit Sicherheit festgestellt werden konnte. Weiterhin wurde untersucht, inwieweit der Regressionskoeffizient b , die Regressionskonstante a sowie das Maß für die Regressionsabweichung e von der Wahl des reduzierten Koeffizientenanzahl P_{red} abhängen. Der Regressionskoeffizient weicht ab einer Mindestgröße von $P_{red} \approx 10^{-2} \cdot P$ für die untersuchten Objekte nur in der Größenordnung 10^{-2} ab. Die Regressionskonstante a schwankt auch bei großen P_{red} in der Größenordnung 10^1 , wird aber für die Skalierungsbestimmung ohnehin nicht berücksichtigt. Das Abweichungsmaß e weicht für große Werte von e ebenfalls in einer Größenordnung von 10^1 ab, sodass $P_{red} > 10^{-1} \cdot P$ zu einem sicheren Erkennen von Skalierungen ausreichen wird. Die Schwelle des Maßes e für eine ausreichend sicher erkannte Skalierung liegt im Bereich $1 \dots 10$. Liegt dieses Maß darunter, kann von einer Skalierung

²¹MPEG-Testset „Lara“, Tests mit unterschiedlichen Texturkompressionsraten, ETH Zürich (Schweiz)

²²Testobjekt „MPI-Dancer“, Max-Planck-Institut für Informatik, Saarbrücken

ausgegangen werden. Im nächsten Schritt ist das Merkmalsmaß w zu bestimmen. Zur Bestimmung des Merkmalsmaßes für Translationen $w_{trans,h}$ war ein Bezug zum Referenzmaß d_{ref} erforderlich, da die Größe der Verschiebungsvektoren durch die Anzahl der Bildpunkte gegeben ist. Dies ist hier nicht erforderlich, weil der Regressionskoeffizient (d. h. der Skalierungsfaktor) bereits ein relatives Maß ist. Das Skalierungsmaß $w_{skal,kontur}$ leitet sich also ausschließlich vom Regressionskoeffizienten b ab. Für $b = 1$ (kein Skalierungsfehler) muss $w_{skal,kontur} \rightarrow 0$ sein. Wenn $b < 1$ (Verkleinerung) muss für $b \rightarrow 0$ das Merkmalsmaß $w_{skal,kontur}$ groß werden. Wenn $b > 1$ (Vergrößerung) muss für ein größer werdendes b das Merkmalsmaß $w_{skal,kontur}$ ebenfalls größer werden. Das erforderliche Verhalten wird von der Logarithmusfunktion erfüllt. Da ein Anstieg des Merkmalsmaßes $w_{skal,kontur}$ sowohl für Verkleinerungen als auch Vergrößerungen gewünscht ist, wird der absolute Betrag des Logarithmus gebildet:

$$w_{skal,kontur} = |\ln b| \quad (4.18)$$

In Tabelle 4.4 werden die Werte von $w_{skal,kontur}$ für verschiedene Fehlertypen vergleichend gegenübergestellt.

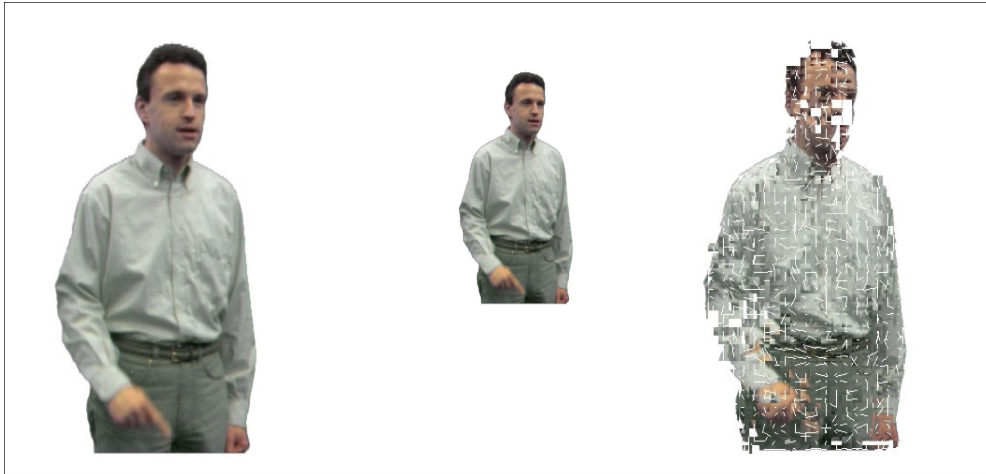


Abbildung 4.9: Referenz-3DVO (links), 3DVO mit Skalierungsfehler (Mitte) und korrigiertes 3DVO (rechts)

Block-Matching kann bei Skalierungsfehlern angewendet werden, wenn für die zu suchenden Blöcke Skalierungstranformationen zugelassen werden. Der Aufwand ist jedoch sehr hoch (siehe [Ohm99b]) und für die Qualitätsbeurteilung nicht unbedingt gerechtfertigt. Es lässt sich jedoch abschätzen, dass das Kriterium des Block-Matching auch ohne Transformationen für viele Skalierungen erfüllt wird. Dies wurde in Tests untersucht. In Abbildung 4.9 ist ein Referenz-3DVO, das zugehörige 3DVO mit einem Skalierungsfehler und die mit Hilfe von hierarchischem Block-Matching korrigierte Ansicht zu sehen. Als Beispiel für eine korrigierte Ansicht wurde die Hierarchieebene 6 gewählt. Zur Verdeutlichung sind die Verschiebungsvektoren dieser Ebene dargestellt (weiß). Die Abbildung beweist, dass hierarchisches Block-Matching bei reinen Skalierungen zu einer Detektion genutzt werden kann. Der nächste Schritt wäre diese Zoom-Bewegung mit einem Modell zu erfassen. Diese Frage wurde für konventionelles Video bereits ausgiebig untersucht [Bon96]. So lassen sich globale Zoom-Bewegungen mit einem 4-Parametermodell vollständig beschreiben, wenn ausreichend viele Blocktranslationen ermittelt wurden. Dies kann man aber nicht für den allgemeinen Fall bei der Qualitätsbewertung voraussetzen.

Beispielsweise führen Bereiche mit ähnlichen Farbeigenschaften zu fehlerhaft ermittelten Verschiebungsvektoren (siehe Abbildung 4.9).

Eine kausale Detektion eines Brennweitenfehlers ist nicht möglich, da die Tiefeninformation nicht bekannt ist. Eine Ermittlung der Tiefeninformation käme dem Prozess der 3DVO-Generierung gleich und würde die Qualitätsbewertung ad absurdum führen. Grundsätzlich ist es aber natürlich möglich, Fehler zu erkennen, die eine ungleichmäßige Skalierung aufweisen. Ein so definierter Fehlertyp lässt sich beispielsweise mit dem Block-Matching ermitteln.

Beliebige Translations-/Skalierungsfehler

Im allgemeinen Fall ist von beliebig gemischten Translations-/Skalierungsfehlern auszugehen. Als Alternativen zur aufwändigen Suche nach einer beliebigen affinen Transformation bieten sich an:

1. Block-Matching und Konturtransformation werden parallel zur Detektion von Translations-/Skalierungsfehlern verwendet. Beide Verfahren erhöhen gegenseitig das Vertrauensmaß der Detektion.
2. Block-Matching und Konturtransformation werden sequentiell angewendet. Die 3DVO-Ansichten werden jeweils korrigiert.

Ein Vorteil von Variante 1 ist das höhere Vertrauensmaß der Detektion. Dagegen spricht jedoch der hohe Aufwand. Insbesondere ist es nur schwer möglich, aus den Verschiebungsvektorfeldern des hierarchischen Block-Matching eine Skalierung zu bestimmen (siehe oben). Ein Vorteil von Variante 2 ist die Korrektur dieser Fehler. Die Korrektur erleichtert die Anwendung nachfolgender Verfahren.

- *Conclusio 4.9:* Zum Beginn der Qualitätsmerkmalsextraktion erfolgt eine sequentielle Detektion von Skalierungs- und Translationsfehlern mit einer anschließenden Korrektur dieser Fehler.

Als Merkmalsmaße zur Erfassung der Translations- und Skalierungsfehler dienen die Maße $w_{trans,BM-h}$, $w_{trans,kontur}$ und $w_{skal,kontur}$. In Tabelle 4.2 sind für verschiedene Fehlertypen die Merkmalsmaße $w_{trans,BM-h}$ der ersten acht Hierarchieebenen aufgelistet. Während $w_{trans,BM-0}$ Translationsfehler der Objektansicht erfasst, können $w_{trans,BM-h}$, ($h > 0$) Verschiebungen von Teilbereichen²³ erfassen.

In Tabelle 4.3 sind die beiden Maße $w_{trans,BM-0}$ und $w_{trans,kontur}$ zur Erkennung von Translationsfehlern für die verschiedenen Fehlertypen gegenübergestellt. Das Maß $w_{trans,BM-0}$ ist wegen der Nutzung der Texturinformation in vielen Fällen genauer, während $w_{trans,kontur}$ wegen der Nutzung der Form die Translation sicherer erkennt.

In Tabelle 4.4 sind für verschiedene Fehlertypen die erkennbaren Skalierungskoeffizienten b und die sich daraus ergebenden Merkmalsmaße $w_{skal,kontur}$ aufgelistet.

4.6.2 Perspektivenfehler des Blickwinkels

Im Gegensatz zu Translations- und Skalierungsfehler wirken sich Änderungen des Blickwinkels²⁴ nicht nur in einer affinen Transformation der Ansicht aus, sondern es sind meist auch andere Teile des Aufnahmeobjektes zu sehen. Dies führt zu mehreren Problemstellungen:

²³z. B. infolge eines Synthese- oder Perspektivenfehlers

²⁴Eine fehlerhafte Objektausrichtung wirkt sich in gleicher Weise aus.

Fehlertyp	$w_{trans,BM-h}$							
	0	1	2	3	4	5	6	7
Translation	0,320	0	0	0	0	0	0	0
Skalierung	0,064	0,087	0,0075	0,0004	0,0008	0,0007	0,0001	0
Verzerrungen	0	0,0010	0,0010	0,0027	0,0007	0,0002	0	0
Translation, Skal.	0,182	0,080	0,0077	0,0008	0,0016	0,0006	0,0001	0
Perspektivenfehler	0,013	0,030	0,015	0,0014	0,0018	0,0006	0	0
zeitl. Fehler	0,0069	0,0057	0,0018	0,0058	0,0004	0,0002	0	0
große Warping-Fehler	0,0311	0,0006	0,0003	0,0005	0,0012	0,0005	0	0
Texturkompression	0,0029	0,0023	0	0,0001	0,0002	0,0001	0	0
Blickwinkel, Transl.	0,096	0,0032	0,0084	0,0045	0,0005	0,0007	0	0
Texturcodierung	0	0	0,0003	0,0001	0,0001	0	0	0
kleine Synthesefehler	0,0017	0,0004	0,0001	0	0	0	0	0

Tabelle 4.2: Merkmalsmaß $w_{trans,BM-h}$ für verschiedene Fehlertypen

Fehlertyp	$w_{trans,BM-0}$	$w_{trans,kontur}$
Translation	0,320	0,367
Skalierung	0,064	0,001
Verzerrungen	0	0,019
Translation, Skalierung	0,182	0,22
Perspektivenfehler	0,013	0,0788
zeitl. Fehler	0,0069	0,047
große Warping-Fehler	0,0311	0,045
Texturkompression	0,0029	0,003
Blickwinkel, Translation	0,096	0,099
Texturcodierung	0	0
kleine Synthesefehler	0,0017	0,012

Tabelle 4.3: Merkmalsmaße $w_{trans,BM-0}$ und $w_{trans,kontur}$ für verschiedene Fehlertypen

- Wenn Teile des Aufnahmeobjektes zu sehen sind, die nicht in der Referenz zu sehen sind, kann der Blickwinkelfehler nicht kausal ermittelt werden. Die Lösung dieses Problems durch eine Bezugnahme zu anderen Ansichten des 3D-Videoobjektes kann im allgemeinen Fall nicht vorgenommen werden.
- Die perspektivischen Änderungen lassen sich nicht kausal zurückführen, da die Tiefeninformation nicht bekannt ist.

An dieser Stelle wird deutlich, dass für die Qualitätsbeurteilung von 3D-Videoobjekten trotz der Nutzung eines Referenz-3DVO nicht für alle Aspekte ein Bezug gegeben ist²⁵. Da eine kausale Rückführung dieses Fehlertyps nicht möglich ist, muss versucht werden, seine spezifischen Auswirkungen zu erfassen. Die Auswirkungen eines fehlerhaften Blickwinkels sind in Abbildung 4.10 dargestellt²⁶. Weiterhin ist die mittels Block-Matching korrigierte 3DVO-Ansicht in Hierarchieebene 2 dargestellt. Es wird deutlich, dass die Ver-

²⁵Daher ist eine Klassifizierung in NR, RR oder FR nicht sinnvoll.

²⁶Testobjekt Nr. 1, J. v. d. Haar, TU Ilmenau [Haa04]

Fehler	b	$w_{skal,kontur}$
Translation	1	0
Skalierung	0,444	1,87
Verzerrungen	nicht erkennbar	0
Translation, Skalierung	0,444	1,87
Perspektivenfehler	nicht erkennbar	0
zeitl. Fehler	nicht erkennbar	0
große Warming-Fehler	nicht erkennbar	0
Texturkompression	1,06	0,13
Blickwinkel, Translation	0,98	0,037
Texturcodierung	1	0
kleine Synthesefehler	nicht erkennbar	0

Tabelle 4.4: Merkmalsmaß $w_{skal,kontur}$ für verschiedene Fehlertypen

schiebungsvektoren nicht als eindeutiges Indiz für einen Blickwinkelfehler dienen können.



Abbildung 4.10: Referenz-3DVO (links), 3DVO mit Blickwinkelfehler (Mitte) und 3DVO mit Block-Matching-Korrektur (rechts)

Untersuchung des örtlichen Informationsgehaltes

Da eine kausale Rückführung nicht möglich ist, sind wiederum die Auswirkungen zu untersuchen. Ein Parameter, welcher zur Bewertung von konventionellem Video genutzt wird, ist die Standardabweichung von SI . Als Beispiel wurde ein 3D-Videoobjekt mit zwei anderen verglichen, welche unter einem abweichenden Blickwinkel aufgenommen wurden (siehe Abbildung 4.11). Um die Differenzen der Standardabweichungen einschätzen zu können, wurden die Standardabweichungen über eine Sequenzdauer von zwei Sekunden gemessen. Erwartungsgemäß liegen die Differenzen der Standardabweichungen für die fehlerhaften Blickwinkel in dem Intervall, wie sie auch temporal variieren.

Nach [Wol02] wird ein blockbezogener Vergleich der Standardabweichungen vorgenommen. Der blockbezogene Vergleich bietet sich bei 3D-Videoobjekten an, um Objektbereiche mit abweichendem Informationsgehalt festzustellen. Dies setzt selbstverständlich eine Korrektur von Translation und Skalierung voraus. Für eine Untersuchung des blockbezogenen Informationsgehaltes wurden die Differenzen der Standardabweichungen des



Abbildung 4.11: Standardabweichungen des örtlichen Informationsgehaltes für Sequenzen mit fehlerhaftem Blickwinkel

örtlichen Informationsgehaltes SI_{stddev} für eine Blockgröße 16×16 bestimmt. Zur Bewertung wurden zunächst das arithmetische Mittel $\overline{\Delta}_{SI_{stddev}}$ der Differenzen (siehe Gleichung 4.19) und der Anteil $a_{2\Delta}$ der mehr als doppelt so großen Differenzen bestimmt (siehe auch [Wol02]).

$$\overline{\Delta}_{SI_{stddev}} = \frac{1}{N} \sum_{i=1}^N |SI_{stddev,ref}(i) - SI_{stddev,eval}(i)| \quad (4.19)$$

In Tabelle 4.5 sind für verschiedene Blickwinkelfehler die Ergebnisse des blockbezogenen Vergleiches aufgelistet. Weiterhin sind diese Messungen für andere Fehlertypen gemessen worden. Die Messwerte zeigen, dass der blockbezogene Vergleich geeignet ist, lokale Fehler zu erkennen, die nicht über eine globale Bestimmung des ΔSI_{stddev} zu bestimmen sind. Es können jedoch damit keine Rückschlüsse auf die Art des Fehlers gezogen werden. Der Anteil der Blöcke mit besonders großen Differenzen ist ebenfalls ein Maß für die Lokalität bestimmter Fehlertypen und es ist ebenso wenig für Rückschlüsse geeignet.

Untersuchung der Farbverteilung

Neben der blockbezogenen Untersuchung des örtlichen Informationsgehaltes kann auch eine blockbezogene Untersuchung der Farbverteilung vorgenommen werden. Hierfür wird der Index $V(H_{ref}, H_{eval})$ des Histogrammvergleiches blockweise bestimmt. Zur Reduktion der zu verwendenden Histogramme wird eine gröbere Quantisierung der Farbkanäle gemäß Gleichung 4.20 vorgenommen. Für die folgenden Messungen wurde $q = 10$ gewählt.

$$M(i, j) = \begin{pmatrix} \lceil r/q \rceil \\ \lceil g/q \rceil \\ \lceil b/q \rceil \end{pmatrix} \Rightarrow \text{inc}(H(r', g', b')) \quad (4.20)$$

Fehlertyp	ΔSI_{stdev}	$\overline{\Delta SI}_{stdev}$	$a_{2\overline{\Delta}}$
Objekt 1, Blickwinkelabweichung 20 Grad	0,45	16,8	31,7%
Objekt 2, Blickwinkelabweichung 20 Grad	2,0	14,1	26,7%
Objekt 3, Blickwinkelabweichung 15 Grad	3,5	11,7	23,0%
Objekt 3, Blickwinkelabweichung 30 Grad	2,93	15,0	20,6%
Objekt 3, Warping-Fehler	1,1	5,6	12,5%
Objekt 3, zeitlicher Fehler 0,8 s	0,0	5,2	14,5%
Objekt 3, zeitlicher Fehler 0,04 s	0,15	4,0	11,8%
Objekt 3, sehr kleine Synthesefehler	0,25	3,0	10,8%
Objekt 3, ohne Fehler	0,0	0,0	0,0%
Objekt 4, Texturkompressionartefakte	0,76	14,5	23,2%
Objekt 5, Texturcodierungsfehler	0,52	8,2	12,4%

Tabelle 4.5: Blockbezogener Vergleich des örtlichen Informationsgehaltes

In Tabelle 4.6 sind die Ergebnisse eines blockbezogenen Histogrammvergleiches für verschiedene Fehlerfälle (vergleiche Tabelle 4.5) aufgelistet. Des Weiteren ist der Index V für die gesamte Objektansicht und der Anteil $a_{0,5\overline{V}}$ der Blöcke mit einem Index V , der kleiner als die Hälfte von \overline{V} (H_{ref}, H_{eval}) ist.

$$\overline{V}(H_{ref}, H_{eval}) = \frac{1}{N} \sum_{i=1}^N V_i(H_{ref}, H_{eval}) \quad (4.21)$$

Fehlertyp	$V(H_{ref}, H_{eval})$	$\overline{V}(H_{ref}, H_{eval})$	$a_{0,5\overline{V}}$
Objekt 1, Blickwinkelabweichung 20 Grad	0,73	0,21	24,7%
Objekt 2, Blickwinkelabweichung 20 Grad	0,51	0,04	7,1%
Objekt 3, Blickwinkelabweichung 15 Grad	0,82	0,20	27,4%
Objekt 3, Blickwinkelabweichung 30 Grad	0,70	0,08	13,2%
Objekt 3, Warping-Fehler	0,89	0,26	20,4%
Objekt 3, zeitlicher Fehler 0,8 s	0,90	0,35	26,1%
Objekt 3, zeitlicher Fehler 0,04 s	0,87	0,209	30,2%
Objekt 3, sehr kleine Synthesefehler	0,92	0,64	20,4%
Objekt 3, ohne Fehler	1,0	1,0	0,0%
Objekt 4, Texturkompressionartefakte	0,65	0,24	45,3%
Objekt 5, Texturcodierungsfehler	0,64	0,08	22,0%

Tabelle 4.6: Blockbezogener Histogrammvergleich

Die Auswertung der Ergebnisse des blockbezogenen Histogrammvergleiches führt zu ähnlichen Schlüssen. Einige Fehlertypen verursachen einen deutlichen Abfall des mittleren blockbezogenen Indexes $\overline{V}(H_{ref}, H_{eval})$ gegenüber dem Gesamtindex, wobei teilweise vermehrt sehr kleine blockbezogene Indizes auftreten.

- *Conclusio 4.10*: Perspektivenfehler des Blickwinkels bewirken eine Änderung statistischer Parameter, z. B. örtlicher Informationsgehalt, Histogramm.
- *Conclusio 4.11*: Die Auswirkungen von Fehlern in bestimmten Bildbereichen lassen sich durch die Anwendung blockbezogener Parameter erkennen.

- *Conclusio 4.12*: Verschiedene Fehlerklassen (Blickwinkelfehler, Warping-Fehler, Texturcodierungsartefakte etc.) bewirken ähnliche Änderungen der statistischen Parameter. Daher ist damit kein Rückschluss auf den Fehlertyp möglich.

Bildung der Merkmalsmaße

Da mit den statistischen Methoden kein Merkmalsmaß w für blickwinkelbedingte Perspektivfehler gebildet werden kann, sind verallgemeinerte Merkmalsmaße entsprechend den verwendeten Methoden zu bestimmen. Die Relation von $\overline{\Delta SI}_{stdev}$ zu ΔSI_{stdev} steht für die Änderungen von SI_{stdev} auf Grund der blockbezogenen Messung²⁷. Dies sind insbesondere globale Veränderungen der Objektansicht, die translations- und skalierungsunabhängig sind (z. B. Änderungen der gesamten Textur, Blickwinkeländerungen). Der Anteil $a_{2\overline{\Delta}}$ steht für die Konzentration dieser veränderten Bildbereiche. Diese Konzentration ist u. a. für kleine Blickwinkelfehler hoch, da sich hier meist nur die Randbereiche und dem Betrachter nahe Objektbereiche ändern. Diese Konzentration lässt sich durch die Wahl der Schwelle s (für die Untersuchungen wurde zunächst $s = 2$ gewählt) anpassen, sodass sich die Konzentration zu $a_{s,\overline{\Delta}}$ ergibt. Mittels der Koeffizienten b_{s1} und b_{s2} erfolgt eine Wichtung dieser beiden Anteile:

$$w_{blockSI} = b_{s1} \cdot \ln \left(\frac{\overline{\Delta SI}_{stdev}}{\Delta SI_{stdev}} \right) + b_{s2} \cdot a_{s,\overline{\Delta}} \quad (4.22)$$

Das Merkmalsmaß $w_{blockSI}$ ist neben den genannten Parametern auch von der Wahl der Blockgröße abhängig. Die Wahl der Mittelung (hier: arithmetisch) und die Bestimmung der Konzentration (hier: Entscheidungsschwelle) können an dieser Stelle noch nicht endgültig festgelegt werden. Das Merkmalsmaß ist an die Fehlerklassen mit gleichen oder ähnlichen (statistischen) Auswirkungen anpassbar.

Die Bildung des Merkmalsmaßes auf Basis des Histogrammvergleiches erfolgt in analoger Weise:

$$w_{blockHI} = b_{h1} \cdot \ln \left(\frac{V(H_{ref}, H_{eval})}{\overline{V}(H_{ref}, H_{eval})} \right) + b_{h2} \cdot a_{k,\overline{V}} \quad (4.23)$$

In Tabelle 4.7 sind für verschiedene Fehlertypen die Merkmalsmaße $w_{blockSI}$ und $w_{blockHI}$ aufgelistet. In einer ersten Näherung wurden $b_{s1} = 0,025$ und $b_{s2} = 1$ sowie $b_{h1} = 0,1$ und $b_{h2} = 1$ gewählt.

4.6.3 Störungen der Silhouette

Eine Veränderung der Silhouette tritt nicht nur bei Perspektivfehlern auf, sondern kann verfahrensspezifische Ursachen haben. Sie können bereits bei der Segmentierung der multiskopischen Aufnahmen oder aber beim Warping, beim Voxel-Colouring etc. entstehen. Aus diesem Grund haben diese Störungen unterschiedliche Charakteristika (vergleiche Abbildung 4.5: Fehler auf Grund einer Texturkompression; Abbildung 4.12: Fehler beim Warping). Die Auswirkungen von Translations- und Skalierungsfehlern wurden in Unterabschnitt 4.6.1 untersucht. Blickwinkelfehler, Objektfehler, zeitliche Fehler etc. bewirken eine nicht rückführbare Änderung der Silhouette.

²⁷Für eine fehlerfreie Ansicht ist $\Delta SI_{stdev} \rightarrow 0$ sowie $\overline{\Delta SI}_{stdev} \rightarrow 0$ und damit $\frac{\overline{\Delta SI}_{stdev}}{\Delta SI_{stdev}} \rightarrow 1$ sowie $\ln \left(\frac{\overline{\Delta SI}_{stdev}}{\Delta SI_{stdev}} \right) \rightarrow 0$.

Fehlertyp	$w_{blockSI}$	$w_{blockHI}$
Objekt 1, Blickwinkelabweichung 20 Grad	0,41	0,37
Objekt 2, Blickwinkelabweichung 20 Grad	0,32	0,32
Objekt 3, Blickwinkelabweichung 15 Grad	0,26	0,42
Objekt 3, Blickwinkelabweichung 30 Grad	0,25	0,35
Objekt 3, Warping-Fehler	0,16	0,33
Objekt 3, zeitlicher Fehler 0,8 s	0,24	0,35
Objekt 3, zeitlicher Fehler 0,04s	0,20	0,44
Objekt 3, sehr kleine Synthesefehler	0,17	0,24
Objekt 3, ohne Fehler	0	0
Objekt 4, Texturkompressionartefakte	0,31	0,55
Objekt 5, Texturcodierungsfehler	0,19	0,43

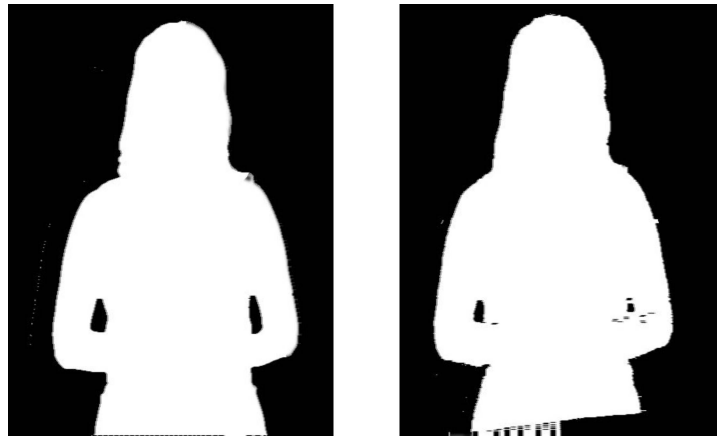
Tabelle 4.7: Merkmalsmaße $w_{blockSI}$ und $w_{blockHI}$ für verschiedene Fehlertypen

Abbildung 4.12: Silhouette eines Referenz-3DVO (links) und eines gestörten 3DVO (rechts)

Konturtransformation

Zunächst wurden die Auswirkungen auf die DFT-Koeffizienten der Kontur untersucht. In Abbildung 4.13 sind die Abweichungen $\Delta |T(i)| = |T_{ref}(i)| - |T_{eval}(i)|$ für $i = 0 \dots 10$ für verschiedene Fehlertypen dargestellt.

Die gemessenen Differenzen $\Delta |T(i)|$ verdeutlichen erwartungsgemäß, dass die Beträge der Koeffizienten durch die verschiedenen Störungen von denen der Referenz abweichen²⁸. Zur Bildung eines Merkmalsmaßes sind zunächst die bereits untersuchten Translations- und Skalierungsfehler auszuschließen. Translationsfehler werden durch die Nichtberücksichtigung von $\Delta |T(0)|$ ausgeschlossen. Skalierungsfehler können ausgeschlossen werden, indem die Koeffizienten $|T_{eval}(i)|$ mit dem ermittelten Skalierungsfaktor b (siehe Unterabschnitt 4.6.1) normiert werden:

$$|T'_{eval}(i)| = \frac{|T_{eval}(i)|}{b} \quad \text{für: } i = 1 \dots P \quad (4.24)$$

Für die nächsten Schritte zur Bildung eines Merkmalsmaßes wird davon ausgegangen, dass die Differenzen $\Delta |T(i)|$ unabhängig von i ein quantitatives Merkmal einer Störung

²⁸Eine Untersuchung der Phasenwinkel wurde nicht vorgenommen, da eine damit ermittelbare Rotation der Kontur als Fehlertyp nicht beobachtet wurde.

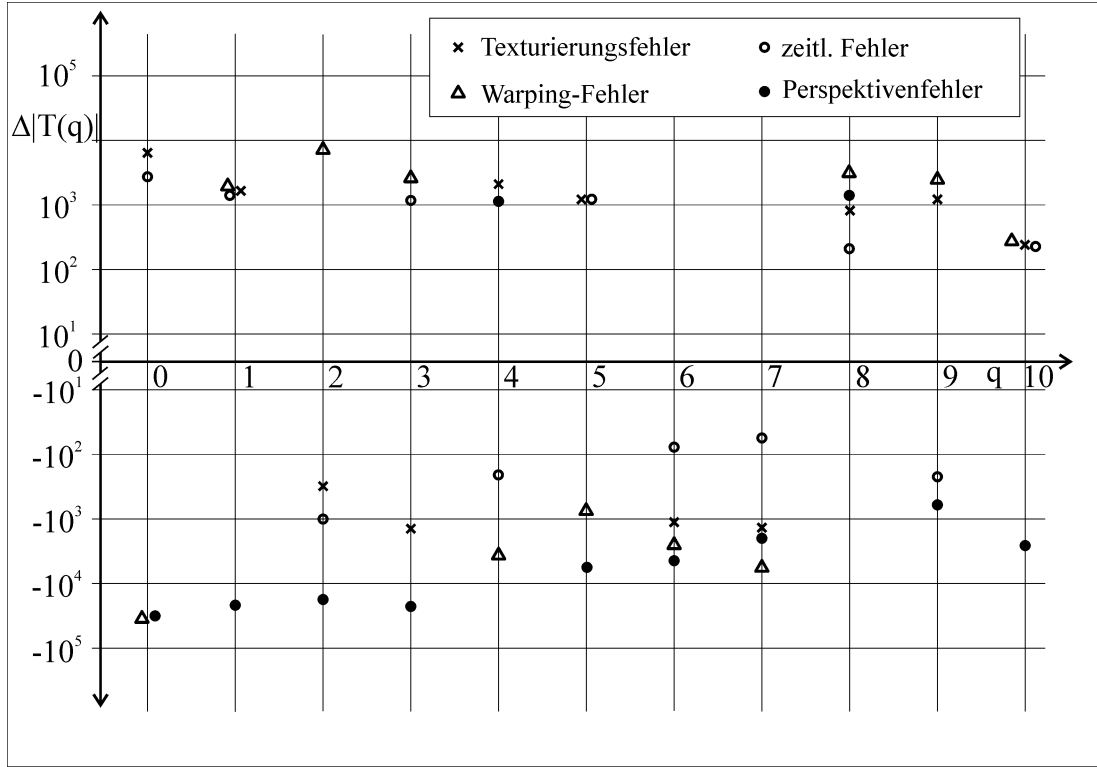


Abbildung 4.13: DFT-Koeffizienten der Kontur für verschiedene Fehlertypen

sind²⁹. Für die Bestimmung des Merkmalsmaßes $w_{silhouette,DFT}$ werden die Differenzen $||T'_{eval}(i)| - |T_{ref}(i)||$ der Koeffizienten bestimmt. Hiermit werden in einem ersten Ansatz niedrig- und hochfrequente Störungen gleich bewertet³⁰.

$$w_{silhouette,DFT} = \frac{1}{P_{red}} \sum_{i=1}^{P_{red}} ||T'_{eval}(i)| - |T_{ref}(i)|| \quad (4.25)$$

- *Conclusio 4.13*: Das Merkmalsmaß $w_{silhouette,DFT}$ auf Basis der Abweichungen der DFT-Koeffizienten eignet sich zur quantitativen Erfassung von Silhouettenstörungen (siehe Tabelle 4.8). Die Ursache (Perspektivenfehler, Synthesefehler etc.) kann damit nicht bestimmt werden.

Im Sinne der Rückführbarkeit der Merkmale wäre eine Unterscheidung von z. B. einem Warping-Fehler und einem Blickwinkelfehler erforderlich. Dies setzt voraus, dass sich die Merkmale in charakteristischen Störungen der Kontur äußern, die sich mittels der spektralen Repräsentation der DFT erfassen lassen. Dies ist allerdings im allgemeinen Fall nicht gegeben. Beispielsweise kann ein Blickwinkelfehler sowohl tieffrequente Änderungen (z. B. Änderungen der Kontur in Abbildung 4.4) als auch hochfrequente Änderungen (z. B. Änderungen der Kontur in Abbildung 4.10). Ebenso können sich Synthesefehler sowohl in tief- als auch hochfrequenten Änderungen auswirken (siehe Abbildung 4.12).

²⁹Dies setzt eine grundsätzliche Ähnlichkeit der beiden Konturen voraus. Diese Annahme ist jedoch keine Einschränkung, da für grundsätzlich verschiedene Konturen Qualitätsuntersuchungen nicht sinnvoll sind.

³⁰Die visuelle Wahrnehmung der Spektralanteile einer Kontur unterscheidet sich von der Wahrnehmung der Ortsspektralanteile, z. B. der zweidimensionalen DCT.

Fehlertyp	$w_{silhouette,DFT}$
Translation	0
Skalierung	15
Konturverzerrungen	933
Translation, Skalierung	15
Perspektiven	1278
zeitl. Konturveränderung 1	315
zeitl. Konturveränderung 2	800
Warping-Fehler 1	518
Warping-Fehler 2	618
keine Störungen	0
Texturkompression	117
Blickwinkel	2448
Texturcodierung	0

Tabelle 4.8: Merkmalsmaß $w_{silhouette,DFT}$ für verschiedene Fehlertypen

Segmentierungsmaße

Für die Bewertung von Objektsegmentierungen werden ebenfalls Masken bzw. Silhouetten untersucht (siehe Unterabschnitt 2.4.4). Nach einer Korrektur von Translations- und Skalierungsfehlern ließen sich die Maße für eine relative Bewertung auch für Silhouetten von 3D-Videoobjekten anwenden.

- Die Maße SQM und D_i^t (siehe Unterabschnitt 2.4.4) basieren auf der Feststellung von falsch klassifizierten Punkten. Daher würden diese Maße bei einer Nutzung für (korrigierte) äußere Silhouetten von 3D-Videoobjekten zu einer negativen Bewertung bei einer Vielzahl von Fehlertypen führen (Blickwinkelfehler, Warping-Fehler etc.). Dies ist derselbe Zusammenhang, wie er bei der Nutzung der Konturtransformation auftritt.
- Die Maße SQM und D_i^t wichten entsprechend dem Abstand der falsch klassifizierten Punkte zum Objekt. Im Falle von 3D-Videoobjekten wäre diese Wichtung nicht sinnvoll (siehe z. B. Abbildung 4.10). Ein spektraler Bezug, wie er bei der Konturtransformation gegeben ist, kommt dem Charakter vieler Fehlertypen näher³¹.
- Das Maß DS^t bewertet die Form auf Basis der Tangentialwinkel und stellt auf diese Weise einen Zusammenhang zur örtlichen Ableitung auf. Damit findet ebenso wie bei der Konturtransformation eine Berücksichtigung des spektralen Charakters von Störungen statt. Zur Bestimmung von DS^t werden die Tangentialwinkel auf der Kontur punktweise bestimmt. Jedoch geht bei 3DVO-Silhouetten der Punktbezug auf der Kontur durch mehrere Störungstypen verloren.

Erfassung von Silhouettenöffnungen

Störungen innerhalb der Silhouette werden mit der Konturtransformation nicht erfasst³². Ein erster Ansatz ist die Bestimmung der Größe B aller korrekten und inkorrekten Öff-

³¹Beipielsweise wird hierdurch eine schmale Warping-Störung auf Basis ihrer Ortsfrequenz und nicht auf Basis ihrer Größe und ihres Abstandes bewertet.

³²Prinzipiell wäre eine Transformation der Kontur von inneren Öffnungen möglich. Damit wird jedoch nicht der Charakter der Störungen erfasst.

nungen i der skalierungskorrigierten Silhouetten. In Tabelle 4.9 werden die verschiedenen Fälle in Abhängigkeit der Größen B der Referenz- und der zu evaluierenden Öffnung unterschieden:

Werte von $B_{eval}(i), B_{ref}(i)$	Ursache
$B_{eval}(i) > 0, B_{ref}(i) = 0$	zusätzliche (falsche) Öffnung
$B_{eval}(i) \neq B_{ref}, B_{eval}(i) > 0, B_{ref}(i) > 0$	korrekte Öffnung, falsche Größe
$B_{eval}(i) = B_{ref}, B_{eval}(i) > 0, B_{ref}(i) > 0$	korrekte Öffnung, korrekte Größe
$B_{eval}(i) = 0, B_{ref}(i) > 0$	verschwundene Öffnung

Tabelle 4.9: Fehlerfälle für Silhouettenöffnungen

Zur Bestimmung der Größen $B(i)$ ist im ersten Schritt eine Zuordnung des Indexes zur Öffnung vorzunehmen. Im allgemeinen Fehlerfall existieren $N_{ref} \neq N_{eval}$ Öffnungen die geometrisch beliebig verteilt sind. Dieser Fall ist jedoch nicht typisch. Ein typisches 3D-Videoobjekt hat $0 \dots 2$ Öffnungen³³. Durch Fehler können weitere kleine Öffnungen im Randbereich entstehen³⁴. Für die Zuordnung des Indexes i können folgende Kriterien angewendet werden:

1. Zuordnung auf Basis des Minimums des euklidischen Abstandes:

$$\min d_{euklid}(A(B_{eval}), A(B_{ref}))$$
2. Zuordnung auf Basis des minimalen Größenunterschiedes: $\min |B_{eval} - B_{ref}|$
3. Minimum der Kriterien aus 1. und 2.

Für die Bildung eines Merkmalsmaßes sind die unterschiedlichen Auswirkungen der Fehlerfälle (Tabelle 4.9) zu berücksichtigen. Insbesondere sind die Fälle zu berücksichtigen, wenn eine Öffnung hinzugefügt wird oder verschwindet. Hierzu werden die Hilfsfunktionen h_{inc} und h_{dec} eingeführt:

$$\begin{aligned} h_{inc}(i) &= 1, \text{ wenn } B_{ref}(i) = 0 \\ h_{inc}(i) &= 0, \text{ sonst} \end{aligned} \quad (4.26)$$

$$\begin{aligned} h_{dec}(i) &= 1, \text{ wenn } B_{eval}(i) = 0 \\ h_{dec}(i) &= 0, \text{ sonst} \end{aligned} \quad (4.27)$$

Für die Bildung der Merkmalsmaße wird unterschieden, ob sich die Anzahl oder die Größe der Öffnungen ändert. Das Merkmalsmodell $w_{\#silhouette}$ erfasst zusätzliche und verschwundene Öffnungen.

$$w_{\#silhouette,num} = \sum_{i=0}^{N_B-1} h_{inc}(i) \cdot B_{eval}(i) + h_{dec}(i) \cdot B_{ref}(i) \quad (4.28)$$

³³Sollten künftig spezielle 3D-Videoobjekte mit vielen Öffnungen erforderlich und entwickelt werden, sind hierfür spezielle Zuordnungsalgorithmen zu schaffen.

³⁴siehe ungestörte Silhouette in Abbildung 4.12

Das Merkmalsmodell $w_{\Delta_{\text{silhouette}}}$ erfasst die Größenunterschiede der Öffnungen. Der Term $(1 - h_{\text{inc}}(i)) \cdot (1 - h_{\text{dec}}(i))$ dient dem Ausschluss der zusätzlichen und verschwundenen Öffnungen.

$$w_{\Delta_{\text{silhouette}}, \text{num}} = \sum_{i=0}^{N_B-1} |B_{\text{ref}}(i) - B_{\text{eval}}(i)| \cdot (1 - h_{\text{inc}}(i)) \cdot (1 - h_{\text{dec}}(i)) \quad (4.29)$$

4.6.4 Fehler bei der Epipolargeometriebestimmung

Eine korrekt gemessene oder korrekt berechnete Epipolargeometrie ist Voraussetzung für viele 3DVO-Verfahren. Die damit ermittelte Fundamentalmatrix \mathbf{F} beschreibt die inneren und äußeren Parameter der Kameras und ist daher für die Vorverarbeitung der Aufnahmen relevant. Die Auswirkungen einer fehlerhaften Fundamentalmatrix wirken sich in den einzelnen 3DVO-Verfahren unterschiedlich aus. Einige Beispiele:

- *Warping*: Zur Durchführung des eindimensionalen Warping findet eine Rektifizierung der Ansichten auf Basis der ermittelten Epipolargeometrie statt. Ist die Rektifizierung fehlerhaft, können die korrekten Korrespondenzen nicht gefunden werden. Das Warping erfolgt entlang der falschen Epipolarlinien.
- *Visual Hull*: Eine fehlerhafte Fundamentalmatrix verzerrt die Visual Hull in jeweils dem Bereich, der durch die betroffenen Ansichten ermittelt wurde. Damit wird dieser Texturbereich ebenso perspektivisch verzerrt.
- *Texturierung von Modellen*: Für die Ermittlung von Texturen für modellbasierte 3DVO-Verfahren werden unterschiedliche Methoden angewendet. Bei der Verwendung eines beliebigen Kamera-Setup werden hierbei in der Berechnung die Zusammenhänge der Epipolargeometrie genutzt. Auch hier führen Fehler zu perspektivischen Verzerrungen von Texturbereichen.

Diese Beispiele verdeutlichen, dass Fehler bei der Epipolargeometriebestimmung sich in verfahrensspezifischen oder allgemeinen Texturierungsfehlern auswirken.

4.6.5 Texturierungsfehler

Eine fehlerhafte Texturierung kann mehrere Ursachen haben, welche in einem hohen Maße mit dem jeweiligen 3DVO-Verfahren verknüpft sind. Die Fehler der Texturen können bereits bei deren Akquisition entstehen, aber auch bei deren Beschreibung (siehe Unterabschnitt 3.3.3) und deren Codierung (siehe Unterabschnitt 4.4.3). Beispiele für deren Auswirkungen sind:

- Ein oder mehrere Teilbereiche des Objektes wirken unscharf oder verwaschen (z. B. Kleidungsteile).
- Sind wesentliche Objektbereiche von den Texturierungsfehlern betroffen (z. B. ein Gesicht), so werden diese sehr deutlich als Verarbeitungsfehler erkennbar.
- Ein oder mehrere Texturbereiche weisen eine falsche Perspektive auf³⁵.

³⁵Dieser Fehlertyp soll durch blickpunktabhängige Texturen und Light-Field-Mapping vermieden werden. Allerdings kann er bereits durch eine mangelhafte Akquisition bedingt sein.

- Das Objekt weist Löcher oder fehlende Randbereiche auf.

Eine Regression der einzelnen Fehlertypen (z. B. Perspektivenfehler eines Texturbereiches) ist auf Grund des Aufwandes und der nur teilweise kausalen Rückführbarkeit nicht durchführbar. Demnach sind wiederum statistische Maße anzuwenden.

Änderungen des örtlichen Informationsgehaltes bei Texturierungsfehlern

In Abbildung 4.14 sind für zwei Objekte mit einem Texturierungsfehler für eine Dauer von zwei Sekunden die Standardabweichungen des örtlichen Informationsgehaltes SI_{stdev} dargestellt. Es handelt sich um Objekte³⁶ mit unterschiedlichen Texturierungsfehlern³⁷. Die hier verwendeten Testobjekte haben synthetischen Charakter³⁸. Da sich natürliche Texturen in ihrem örtlichen Informationsgehalt von synthetischen Texturen unterscheiden, werden sich für natürliche 3D-Videoobjekte³⁹ andere typische Werte für SI_{stdev} ergeben. Die Fehlerauswirkungen werden sich jedoch prinzipiell nicht unterscheiden.

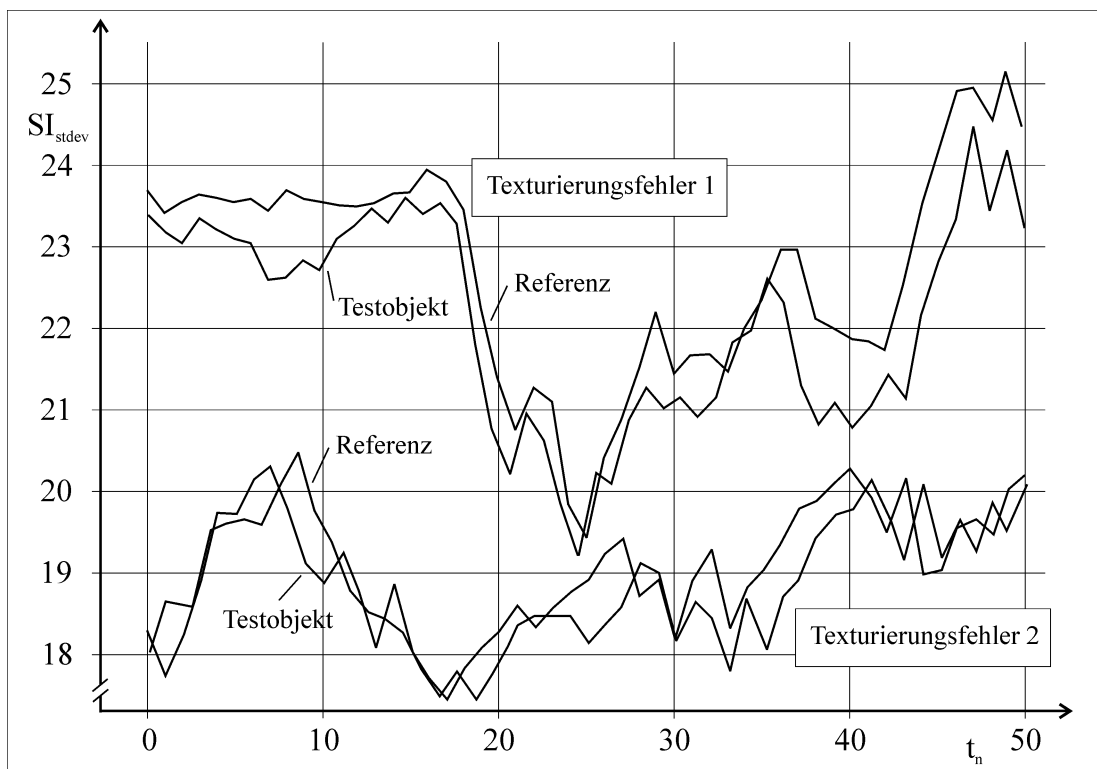


Abbildung 4.14: Örtl. Informationsgehalt für zwei Sequenzen mit Texturierungsfehlern

Zur Bildung eines Merkmalsmodells wird die Differenz der Standardabweichungen von SI bestimmt.

$$w_{SI} = |SI_{stdev,ref} - SI_{stdev,eval}| \quad (4.30)$$

In Tabelle 4.10 sind für verschiedene Fehlerfälle die Werte für w_{SI} angegeben. Hiermit wird deutlich, dass dieses statistische Maß für Texturierungsfehler geeignet ist, welche den örtlichen Informationsgehalt signifikant beeinflussen. Dies ist insbesondere im Falle der Kompression von Texturdaten gegeben.

³⁶1: „MPI-Dancer“, Max-Planck-Institut Saarbrücken; 2: „Lara“, ETH Zürich (Schweiz)

³⁷Texturierung,/-codierung, Kompression

³⁸Das Objekt „Lara“ ist ein rein synthetisches Objekt.

³⁹Natürliche modellbasierte 3D-Videoobjekte befinden sich in einem frühen Entwicklungsstadium.

Fehlertyp	w_{SI}
Translation	0
keine Störungen	0
Verzerrungen im Konturbereich	0,173
Perspektivenfehler	2,47
zeitl. Fehler 0,4 s	1,98
zeitl. Fehler 0,04 s	0,29
Warping-Fehler	2,1
Störungen durch Texturkompression 1	11,9
Störungen durch Texturkompression 2	12,6
Blickwinkelfehler 10 Grad	0,144
Blickwinkelfehler 20 Grad	4,60
Texturcodierungsartefakte	2,39

Tabelle 4.10: Merkmalsmaß w_{SI} für verschiedene Fehlertypen

Änderungen der Farbverteilung bei Texturierungsfehlern

Als weitere statistische Methode wurde wiederum ein Histogrammvergleich vorgenommen. In Abbildung 4.15 ist der Index $V(H_{ref}, H_{eval})$ des Histogrammvergleiches⁴⁰ für zwei Objekte mit einem Texturierungsfehler (siehe oben) für eine Dauer von zwei Sekunden dargestellt. Erwartungsgemäß sinkt der Index und schwankt um einen Mittelwert.

Für die Bildung des Merkmalsmaßes w_{HI} wird der reziproke Wert des Histogrammvergleichsindexes bestimmt. Zur Berücksichtigung des Verhaltens von $V(H_{ref}, H_{eval})^{-1} \rightarrow 0$ und $V(H_{ref}, H_{eval})^{-1} \rightarrow 1$ wird der natürliche Logarithmus gebildet.

$$w_{HI} = \ln \frac{1}{V(H_{ref}, H_{eval})} \quad (4.31)$$

In Tabelle 4.11 sind für verschiedene Fehlerfälle die Werte für w_{HI} angegeben. Selbstverständlich sinkt der Histogrammvergleichsindex nicht nur bei den untersuchten Texturierungsfehlern, sodass sich ein erhöhtes w_{HI} beispielsweise auch für Perspektivenfehler ergibt.

Fehlertyp	w_{HI}
Translation	0
keine Störungen	0
Verzerrungen im Konturbereich	0,11
Perspektivenfehler	1,56
zeitl. Fehler 0,04 s	0,32
Warping-Fehler	0,79
Störungen durch Texturkompression 1	1,7
Störungen durch Texturkompression 2	1,0
Blickwinkelfehler 10 Grad	0,73
Texturcodierungsartefakte	1,0

Tabelle 4.11: Merkmalsmaß w_{HI} für verschiedene Fehlertypen

⁴⁰entsprechend der Gleichung 4.20



Abbildung 4.15: Verlauf des Histogrammvergleichsindexes für zwei Sequenzen mit Texturierungsfehlern

4.6.6 Verdeckungsfehler

Die einzelnen 3DVO-Verfahren behandeln Okklusionen in unterschiedlicher Weise. Oftmals wird der verdeckte Bereich mit dem Bildinhalt einer benachbarten Aufnahme gefüllt. Teilweise führen Okklusionen aber auch zu Löchern im 3D-Videoobjekt. Die Beeinträchtigung ist natürlich immer dann hoch, wenn relevante Objektteile (z. B. Gesicht, Hände) betroffen sind.

Ebenso wie bei den Texturierungsfehlern lassen die unterschiedlichen Auswirkungen eine Rückführung von Okklusionen nicht zu. Daher sind die Auswirkungen auf statistische Maße und andere Merkmalsmaße zu untersuchen:

- *Füllung durch benachbarte Textur*: Änderung von statistischen Maßen (z. B. SI_{stdev} , $V(H_{ref}, H_{eval})$)
- *Füllung mit konstanten Farbwert*: Änderung von statistischen Maßen (z. B. SI_{stdev} , $V(H_{ref}, H_{eval})$), jedoch größere Änderung als bei einer Füllung mittels Textur
- *keine Füllung (Öffnung)*: Erfassung der zusätzlichen Öffnung (Loch) mittels $w_{\#silhouette,num}$

Diese Fallunterscheidung zeigt, dass obwohl keine Rückführung von Okklusionen durchgeführt wird, eine Erfassung der (visuellen) Auswirkung möglich ist. Daher ergibt sich kein weiteres Merkmalsmaß $w_{\Psi, \chi}$.

4.6.7 Weitere Merkmalsmodelle

Für die Methodik zur Qualitätsbeurteilung von 3D-Videoobjekten und zur Bildung eines zugehörigen Qualitätsmaßes ist es unerlässlich, dass weitere Merkmalsmodelle berücksichtigt werden können.

1. Die in dieser Arbeit untersuchten und verwendeten mathematischen Methoden bilden eine begrenzte Auswahl. Es ist davon auszugehen, dass weitere Methoden χ zur Anwendung kommen bzw. entwickelt werden. Mit diesen Methoden χ können weitere Merkmalsmodelle geschaffen werden, welche u. a. eine genauere Rückführung zu den 3DVO-Eigenschaften ermöglichen.
2. Neu- und Weiterentwicklungen von 3DVO-Verfahren führen zu Objekteigenschaften (Qualitätsmerkmalen Ψ), für die neue Merkmalsmodelle $w_{\Psi,\chi}$ erforderlich sind.

4.7 Dynamik der Merkmalsmaße

In diesem Abschnitt werden die dynamischen Eigenschaften der Merkmalsmaße $w_{\Psi,\chi}$, w_χ untersucht. Diese Eigenschaften führen zu Merkmalsmaßen des gesamten Videoobjektes der Dauer T bzw. N . Im Unterabschnitt 4.5.3 wurde ein Ansatz zur Untersuchung der zeitabhängigen und blickpunktsänderungsabhängigen Merkmale vorgestellt. Im nächsten Unterabschnitt werden Modelle zur Erfassung der Zeitabhängigkeit der Merkmalsmaße vorgestellt.

4.7.1 Zeitabhängigkeit der Merkmalsmaße

Charakteristika

Die zeitliche Abhängigkeit der verschiedenen Merkmalsmodelle weist unterschiedliche Charakteristika auf:

- *stationär, deterministisch*: Das zurückgeführte Merkmal hat einen zeitlich konstanten Charakter: Translationen, Skalierungen, zusätzliche/fehlende Silhouettenöffnungen
- *variabel, deterministisch*: Das zurückgeführte Merkmal hat einen zeitlich veränderlichen Charakter: veränderliche Translationen, Skalierungen usw.
- *stationär, statistisch*: Das statistische Merkmal hat einen zeitlich konstanten Charakter: w_{SI} , w_{HI} usw.
- *variabel, statistisch*: Das statistische Merkmal hat einen zeitlich veränderlichen Charakter: veränderliches w_{SI} , w_{HI} usw.

Die genannten Charakteristika unterscheiden sich zum Teil deutlich in ihrer Qualitätsbeeinflussung. Beispielsweise wird eine permanente geringfügige Verkleinerung eines Objektes (stationär, deterministisch) die Qualität kaum oder gar nicht beeinträchtigen. Im Gegensatz dazu ist eine veränderliche Translation (variabel, deterministisch) sehr störend. Bei den statistischen Merkmalen kann man nicht auf das Zeitverhalten eines einzelnen Fehlers schließen, da mehrere Fehlertypen das Merkmal beeinflussen können.

Zeitliches Mittel, zeitliche Abweichung

Für einige Merkmale (z. B. w_{SI}) ist das zeitliche Mittel eine signifikante Eigenschaft. Für ein Merkmal $w_{\Psi,\chi}$ sei $wm_{\Psi,\chi}$ das zeitliche Mittel. Hierfür werden alle $w_{\Psi,\chi}(i)$ einer jeden Ansicht i einer Sequenz der Länge N berücksichtigt.

$$wm_{\Psi,\chi} = \frac{1}{N} \sum_{i=0}^{N-1} w_{\Psi,\chi}(i) \quad (4.32)$$

Wenn ein Merkmalsmaß zeitlich variabel ist, sei diese Eigenschaft durch das Maß $ws_{\Psi,\chi}$ beschrieben. Hierfür wird die Standardabweichung von $w_{\Psi,\chi}(i)$ genutzt.

$$ws_{\Psi,\chi} = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (w_{\Psi,\chi}(i) - \bar{w}_{\Psi,\chi})^2} \quad (4.33)$$

Erfassung von kurzzeitigen Merkmalen

Selbstverständlich gibt es häufig Störungen, welche nur kurzzeitig auftreten. Beispielsweise kann ein 3D-Videoobjekt während einer Navigation von verschiedenen Blickpunkten stark gestört sein. Im zeitlichen Mittel sind diese Fehler klein, jedoch müssen sie für die Qualitätsbewertung erfasst werden. Eine simple Verwendung der Spitzenwerte eines Merkmalsmaßes während der Objektdauer T ist jedoch kaum geeignet. Dieser Spitzenwert könnte beispielsweise von einem für eine einzelne Ansicht falsch bestimmten Merkmalsmaß herrühren. Eine geeignetere Methode ist die Bestimmung des Anteiles $a_{c,w}$ der Merkmalsmaße, welche über einer bestimmten Schwelle liegen⁴¹. Diese Schwelle kann mit dem c -fachen des Mittelwertes $\bar{w}_{\Psi,\chi}$ z. B. mit $c = 2$ angesetzt werden. Der Anteil $a_{c,w}$ dient als Koeffizient für den mittleren Momentanwert.

$$wa_{\Psi,\chi} = a_{c,w} \cdot \bar{w}_{\Psi,\chi} \quad (4.34)$$

Beispiele für dynamische Merkmalsmaße

In Tabelle 4.12 sind einige Beispiele für die dynamischen Merkmalsmaße $wm_{\Psi,\chi}$, $ws_{\Psi,\chi}$ und $wa_{\Psi,\chi}$ erläutert.

Merkmalsmaß	Erläuterung
$wm_{trans-0,BM}, wm_{trans,kontur}$	gesamter Translationsfehler während einer Sequenz
$wa_{trans-0,BM}, wa_{trans,kontur}$	kurzzeitige Translationsfehler
$wm_{\#silhouette,num}$	zusätzliche/fehlende Silhouettenöffnungen
$ws_{\#silhouette,num}$	Auftreten/Verschwinden von Silhouettenöffnungen
$wm_{\Delta silhouette,num}$	fehlerhafte Größe von Silhouettenöffnungen
$ws_{\Delta silhouette,num}$	Schwankung der Größenfehler der Silhouettenöffnungen
wm_{SI}	Gesamtgröße des SI -basierten Merkmals
ws_{SI}	Änderung des SI -basierten Merkmals

Tabelle 4.12: Beispiele für dynamische Merkmalsmaße

In Tabelle 4.13 sind für die folgenden Testobjekte einige dynamische Merkmalsmaße gegenübergestellt.

⁴¹Diese Methode hat sich für die Bildung von Videoqualitätsmaßen bewährt [Wol02].

1. 3D-Videotestobjekt 1 (siehe Abbildung 4.6) mit minimalen Warping-Fehlern, minimalen Perspektivenfehlern und einigen Silhouettenfehlern, Dauer: 3,6 s
2. 3D-Videotestobjekt 1 mit wenigen Warping-Fehlern, minimalen Perspektivenfehlern und einigen Silhouettenfehlern, Dauer: 3,6 s
3. 3D-Videotestobjekt 1 mit deutlichen Warping-Fehlern, minimalen Perspektivenfehlern und einigen Silhouettenfehlern, Dauer: 3,6 s
4. 3D-Videotestobjekt 1 mit kurzzeitigen starken Warping-Fehlern, kurzzeitigen Perspektivenfehlern und einigen Silhouettenfehlern, Dauer: 3,6 s
5. 3D-Videotestobjekt 1 ohne Störungen, Dauer: 3,6 s
6. 3D-Videotestobjekt 2 (siehe Abbildung 4.5) mit Texturkompressionsartefakten, Dauer: 2 s
7. 3D-Videotestobjekt 3 (siehe Abbildung 4.7) mit Texturcodierungsartefakten, Dauer: 2 s

Merkmalsmaß	Testobjekt						
	1	2	3	4	5	6	7
$wm_{trans-0,BM}$	0	0,00062	0,0011	0,0080	0	0,0022	0
$ws_{trans-0,BM}$	0	0,00083	0,00089	0,012	0	0,0012	0
$wa_{trans-0,BM}$	0	0,00022	0,00061	0,0020	0	0,00004	0
$wm_{silhouette,DFT}$	1110	930	1360	990	0	240	0
$ws_{silhouette,DFT}$	1800	910	1220	1100	0	130	0
$ws_{silhouette,DFT}$	0,122	0,078	0,078	0,17	0	0,059	0
wm_{SI}	0,163	0,155	0,351	1,72	0	0,368	0,511
ws_{SI}	0,182	0,138	0,263	2,745	0	0,295	0,310
wa_{SI}	0,0235	0,0223	0,0263	0,381	0	0,0360	0,050

Tabelle 4.13: Werte der dynamische Merkmalsmaße für verschiedene Testobjekte

4.7.2 Blickpunktsänderungsabhängigkeit

Wie im Unterabschnitt 4.5.3 erläutert wurde, können Aussagen über die Blickpunktsänderungsabhängigkeit getroffen werden, wenn die dynamischen Merkmalsmaße $wm_{\Psi,\chi}$, $ws_{\Psi,\chi}$, $wa_{\Psi,\chi}$ auf den bekannten Navigationspfad bezogen werden. Hierfür muss zu jedem Zeitpunkt T , d. h. für jede Ansicht i , der Betrachterstandpunkt $\mathbf{P}(i)$, die Blickrichtung $\mathbf{D}(i)$ ⁴² und der Sehbereich⁴³ $\phi(i)$ ⁴⁴ bekannt sein⁴⁵. Weiterhin muss der Ort des Objektes bekannt

⁴²Der Rotationswinkel $\omega_D(i)$ auf der Blickrichtungsachse ist für viele Anwendungen konstant und wird daher mittels $\mathbf{D}'(i)$ zunächst nicht berücksichtigt.

⁴³Äquivalent für Brennweite, Seitenverhältnis; *field of view*

⁴⁴Bei festen Seitenverhältnis genügt der horizontale bzw. diagonale Öffnungswinkel.

⁴⁵In der Computergrafik-/animation gibt es weitere mathematische Formen, einen Betrachter (bzw. eine virtuelle Kamera) zu beschreiben. Die hier verwendete Form entspricht z. B. VRML.

sein. Er wird für die weiteren Betrachtungen in den Ursprung des Koordinatensystems gesetzt.

$$\mathbf{P}(i) = \begin{bmatrix} x_P(i) \\ y_P(i) \\ z_P(i) \end{bmatrix} \quad \mathbf{D}(i) = \begin{bmatrix} x_D(i) \\ y_D(i) \\ z_D(i) \\ \omega_D(i) \end{bmatrix} \quad \mathbf{D}'(i) = \begin{bmatrix} x_D(i) \\ y_D(i) \\ z_D(i) \end{bmatrix} \quad (4.35)$$

Die visuelle Wahrnehmung hängt vom Charakter der Navigation ab. So werden Störungen intensiver wahrgenommen, wenn man sich beispielsweise auf ein Objekt zu bewegt, als wenn man sich beliebig durch den virtuellen Raum bewegt. Eine genaue Berücksichtigung des Navigationscharakters erfordert Erfahrungen und Tests des Nutzungsverhaltens. Daher wird hier auf der Grundlage bestehender Erfahrungen zunächst folgende Unterscheidung getroffen:

1. Der Betrachter bewegt sich geradlinig auf das Objekt zu oder weg und blickt dabei auf das Objekt.
2. Der Betrachter bewegt sich mit um das Objekt herum und blickt dabei auf das Objekt.
3. Der Betrachter bewegt sich beliebig im Raum und blickt dabei in beliebige Richtungen.
4. Der Betrachter bewegt sich nicht und blickt auf das Objekt.

Diese Navigationstypen lassen sich wie folgt erkennen:

1. Die Blickrichtung bleibt nahezu konstant auf das Objekt $\mathbf{D}'(i) = c \cdot (x_P, y_P, z_P)^T$. Der Abstand $r(i) = \sqrt{x_P^2(i) + y_P^2(i) + z_P^2(i)}$ ändert sich stetig.
2. Der Abstand $r(i) = \sqrt{x_P^2(i) + y_P^2(i) + z_P^2(i)}$ ändert sich wenig oder nicht. Die Blickrichtung ist permanent auf das Objekt im Ursprung, d. h. $\mathbf{D}'(i) = \mathbf{P}(i)$.
3. Standpunkt $\mathbf{P}(i)$, Blickrichtung $\mathbf{D}'(i)$ und Abstand $r(i)$ ändern sich beliebig.
4. Die Blickrichtung ist permanent auf das Objekt im Ursprung, d. h. $\mathbf{D}'(i) = \mathbf{P}(i)$. Der Standpunkt ändert sich nicht.

Zur Erfassung werden vier Parameter definiert. Der Parameter nav_v beschreibt die Geschwindigkeit, der Parameter nav_{um} beschreibt Navigationen um das Objekt und der Parameter nav_{obj} Bewegungen auf oder vom Objekt weg. Der Parameter nav_{still} beschreibt eine stillstehende Betrachtung des Objektes.

$$nav_v = \frac{1}{m_2 - m_1} \sum_{i=m_1}^{m_2-1} r(i) - r(i+1) \quad (4.36)$$

$$nav_{um} = nav_v, \text{ wenn:}$$

$$r(i) \approx \bar{r} \text{ und}$$

$$\mathbf{D}'(i) \approx \mathbf{P}(i) \text{ für alle } m_1 < i < m_2, \text{ sonst}$$

$$nav_{um} = 0 \quad (4.37)$$

$$nav_{obj} = nav_v, \text{ wenn}$$

$$\begin{aligned}
& \mathbf{D}'(i) \approx \mathbf{P}(i) = c \cdot \mathbf{P}(i-1) \text{ für alle } m_1 < i < m_2, \text{ sonst} \\
nav_{obj} &= 0 \\
nav_{still} &= 1, \text{ wenn}
\end{aligned} \tag{4.38}$$

$$\begin{aligned}
& \mathbf{D}'(i) \approx \mathbf{P}(i) = \mathbf{P}(i-1) \text{ für alle } m_1 < i < m_2 \text{ und} \\
& nav_v \approx 0, \text{ sonst} \\
nav_{still} &= 0
\end{aligned} \tag{4.39}$$

Die Parameter können zur Einschätzung des Charakters einer gesamten Navigation ($m_1 = 0, m_2 = N$) oder eines Teiles $m_1 \dots m_2$ genutzt werden. Dies wird im folgenden Beispiel verdeutlicht:

- *Beispiel:* Ein 3D-Videoobjekt (siehe Abbildung 3.1) wurde für die Verwendung innerhalb eines synthetischen 3D-Szenarios generiert. In einem Anwendungsfall bewegt sich der Nutzer in einem Winkel von ca. 60 Grad um das Objekt, hält kurz inne und danach bewegt er sich etwas schneller wieder zurück zum Ausgangspunkt. Der Nutzer blickt während der gesamten Navigation auf das Objekt. Die Parameter ergeben sich zu:

Phase 1: $nav_v = 6,50$; $nav_{um} = 6,50$; $nav_{obj} = 0$; $nav_{still} = 0$

Phase 2: $nav_v = 0$; $nav_{um} = 0$; $nav_{obj} = 0$; $nav_{still} = 1$

Phase 3: $nav_v = 6,35$; $nav_{um} = 6,35$; $nav_{obj} = 0$; $nav_{still} = 0$

5 3D-Videoobjektqualitätsmaß: 3DVQM

Im vorherigen Kapitel wurden ein Modell für die Qualität von 3D-Videoobjekten sowie einige Maße für die festgestellten Qualitätsmerkmale entwickelt. Es bedarf nun einer Methodik, mit Hilfe dieser Merkmalsmaße Qualitätsparameter zu definieren, welche letztlich zu einem 3D-Videoobjektqualitätsmaß führen.

5.1 Methodik für die Bestimmung der Qualitätsparameter

5.1.1 Anwendung der Merkmalsmodelle

Die in den Abschnitten 4.6 und 4.7 beschriebenen Merkmalsmodelle zeigen:

- *Conclusio 5.1*: Nur ein Teil der Qualitätsmerkmale lassen sich durch ein Modell kausal zurückführen.
- *Conclusio 5.2*: Einige Maße sind von mehreren Qualitätsmerkmalen abhängig.
- *Conclusio 5.3*: Einige Modelle erfordern eine Korrektur von Translations- und Skalierungsfehlern.

Die erste und die zweite Conclusio zeigen, dass die Anwendung eines Faktorenmodells an Grenzen stoßen wird. Aus der dritten Conclusio folgt, dass der erste Schritt der zu entwickelnden Methode die Erkennung und die Korrektur von Translations- und Skalierungsfehlern sein muss.

5.1.2 Abfolge zur Bestimmung der Merkmalsmaße und Qualitätsparameter

Aus den entwickelten Merkmalsmodellen und den oben genannten Folgerungen lässt sich unmittelbar eine Abfolge zur Bestimmung der Merkmalsmaße und Qualitätsparameter ableiten. In Abbildung 5.1 ist diese Abfolge bis zur Bildung eines 3D-Videoobjektqualitätsmaßes dargestellt.

Ein wichtiges Kriterium für diese Methodik ist die Eignung für Weiterentwicklungen und Erweiterungen:

- Die Korrektur von Translations- und Skalierungsfehlern erlaubt in einem begrenzten Umfang die Nutzung bildpunktbasierter Methoden. Gleichzeitig kann aber jede beliebige bildpunktsunabhängige Methode genutzt werden.
- Die Merkmalsmodelle können beliebig erweitert und verbessert werden. Dies ist wichtig, da Neu- und Weiterentwicklungen von 3DVO-Verfahren zu erwarten sind.
- Die Qualitätsparameter und deren Zusammenführung können beliebig angepasst und verbessert werden. So können künftige Erfahrungen aus Applikationen und subjektiven Tests in die Bildung des 3D-Videoobjektqualitätsmaßes einfließen.

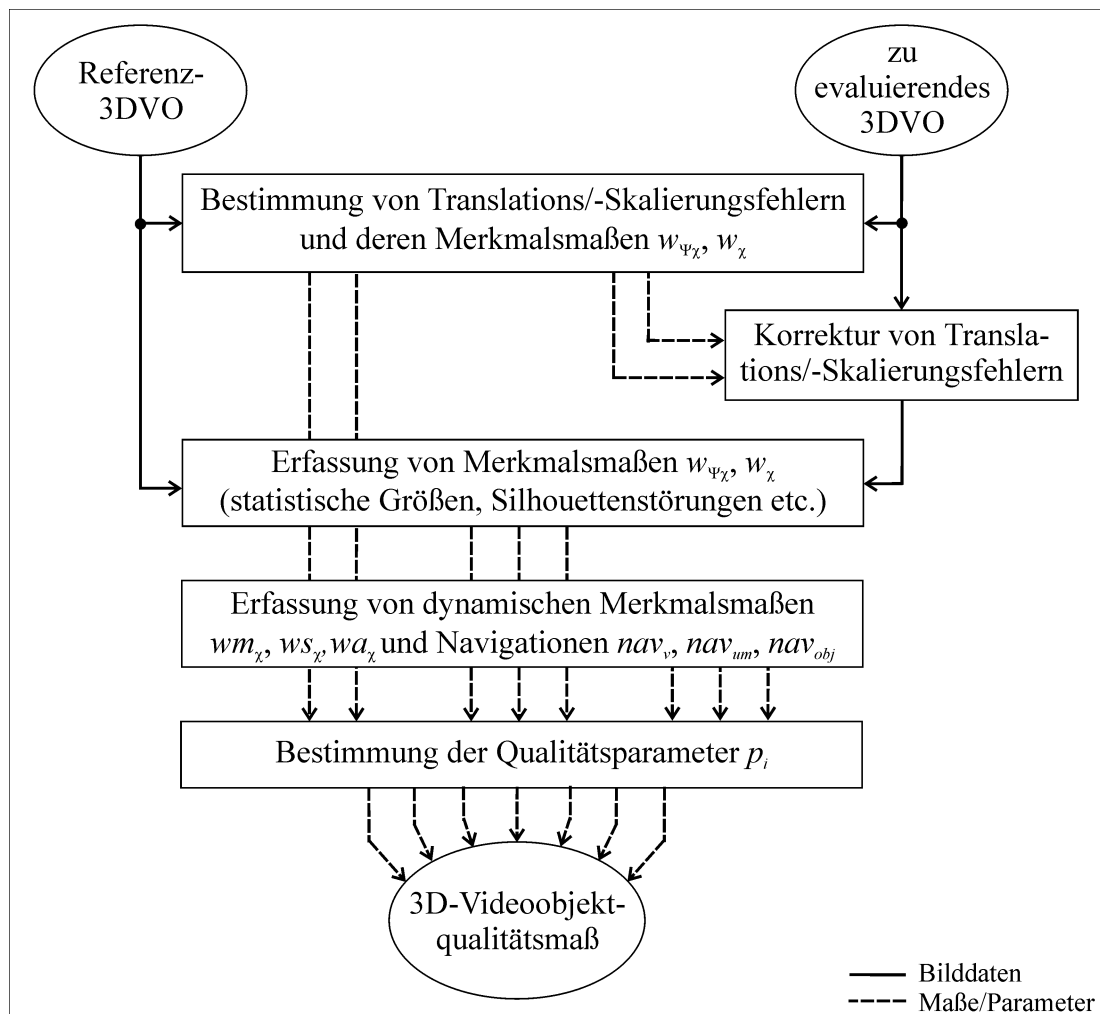


Abbildung 5.1: Abfolge zur Bestimmung des 3D-Videoobjektqualitätsmaßes

5.1.3 Regression von Qualitätsparametern

Nutzung von Faktorenmodellen

Die Qualitätsparameter p_i sollen dazu dienen, Zusammenhänge zwischen den Merkmalsmaßen und deren Qualitätsbeeinflussung aufzustellen. Solche Zusammenhänge wurden in anderen Arbeiten auf unterschiedliche Weise beschrieben. Viele Faktorenmodelle nutzen hierfür Gewichtungsfaktoren, welche beispielsweise mit Hilfe der Cluster-/Faktorenanalyse ermittelt wurden [Kay96], [Fec00]¹. Die Anpassung mittels Gewichtungsfaktoren hat den Nachteil, dass damit häufig nur in einem Teil des Wertebereiches der subjektiven Wahrnehmung entsprochen wird. In [Wol02] erfolgt die Anpassung in einem ersten Schritt mittels diverser Methoden (u. a. Vergleichsfunktionen, Skalierungen; siehe Unterabschnitt 2.4.2). Im zweiten Schritt erfolgt die Gewichtung mittels Faktoren. Durch diese Anpassung erreicht man mit diesem Faktorenmodell eine sehr gute Annäherung an die subjektive Wahrnehmung [VQEG03].

¹In [Fec00] werden mehrere Faktorenmodelle vergleichend vorgestellt, welche ebenfalls Gewichtungsfaktoren verwenden.

Ansatz für die Ableitung der Qualitätsparameter

Für die Bewertung von 3D-Videoobjekten muss berücksichtigt werden, dass die entwickelten Merkmalsmodelle nur eine teilweise kausale Rückführung ermöglichen. Weiterhin ist zu beachten, dass einige Merkmalsmodelle von mehreren Störungstypen abhängig sind².

- *Conclusio 5.4*: Die Qualitätsparameter für 3D-Videoobjekte bilden das Bindeglied zwischen den subjektiven Qualitätsmerkmalen (Störungen) und den objektiv erfassbaren Merkmalen.

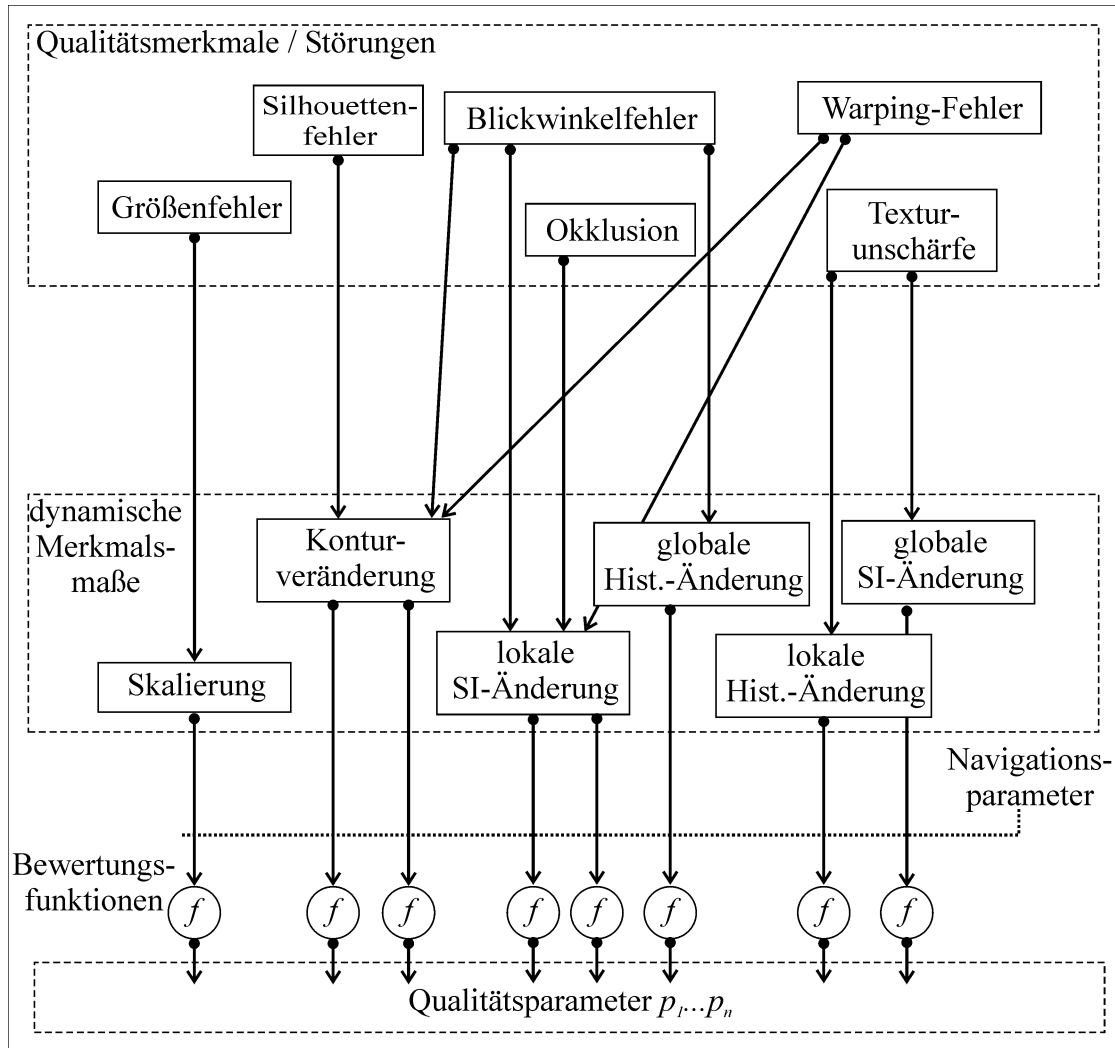


Abbildung 5.2: Bildung der Qualitätsparameter (Beispiele)

In Abbildung 5.2 ist der Zusammenhang zwischen Qualitätsmerkmalen, objektiven Merkmalsmodellen und Qualitätsparametern für einige Beispiele dargestellt. Zur Bildung der Qualitätsparameter werden nur die dynamischen Merkmalsmaße genutzt, da nur diese einen Bezug zur gesamten Sequenz (d. h. zum gesamten Videoobjekt) aufweisen. Selbstverständlich können aus einem Merkmalsmaß auch mehrere Qualitätsparameter abgeleitet

²Dies gilt ebenfalls für einige Merkmalsfaktoren von konventionellem Video, da z. B. die Änderung eines SI-basierten Merkmals mehrere Störungstypen als Ursache haben kann. Diese Problemstellung egalisiert sich teilweise, da der Einfluss der verschiedenen Störungstypen auf die subjektive Wahrnehmung wiederum auch ähnlich ist.

werden, um beispielsweise den Einfluss der Navigation mit einem weiteren Parameter zu erfassen.

Bewertungsfunktionen für 3DVO-Qualitätsparameter

Für die Ermittlung von Qualitätsparametern aus den Merkmalsmaßen sind neben linearen und nichtlinearen Funktionen auch Vergleichsfunktionen, Abstandsmaße etc. erforderlich. Eine Vielzahl von Bewertungsfunktionen wird in [Wol02] vorgestellt und in verifizierten Qualitätsmaßen angewendet. Für die Bestimmung der 3D-Videoobjektqualitätsparameter kommt folgende Auswahl zur Anwendung:

- *nonlineare Skalierung*: Modellierung eines exponentiellen Einflusses
- *Begrenzung des Wertebereiches (Clipping)*: Erfassung des Einflusses über/unter bestimmten Schwellenwerten
- *Abstandsmaße*: Abhängigkeit zwischen Merkmalsmaßen

Mit Hilfe dieser Bewertungsfunktionen werden die Qualitätsparameter p_ψ gebildet. Ein Qualitätsparameter p_ψ erfasst das Modell ψ eines Qualitätsaspektes und ist abhängig von einem oder mehreren Merkmalsmaßen $w_{\Psi,\chi}$.

5.1.4 Exemplarische Bildung einiger Qualitätsparameter

Mit Hilfe der in Kapitel 4 entwickelten Merkmalsmaße wird die Bildung einer Auswahl an Qualitätsparametern vorgenommen.

Positionsveränderungen

Im Abschnitt 4.6 wurden die Merkmalsmaße $w_{trans,BM-0}$ und $w_{trans,kontur}$ zur Ermittlung von Translationsfehlern vorgestellt. Da beide Methoden (Block-Matching und Konturtransformation) beim Einwirken weiterer Fehler die Translationsstörung vermindert gut repräsentieren³, wird der reziproke quadratische Abstand beider Maße als Wichtungsfaktor für das mittlere Maß genutzt.

$$\overline{w}_{trans,\chi} = \frac{w_{trans,BM-0} + w_{trans,kontur}}{2 \cdot ((w_{trans,BM-0} - w_{trans,kontur})^2 + 1)} \quad (5.1)$$

Die drei abgeleiteten dynamischen Maße entsprechen den drei zu berücksichtigenden Aspekten von Translationsfehlern:

1. Das zeitliche Mittel $w_{m_{trans,\chi}}$ entspricht einem permanent vorhandenen Translationsfehler. Sein Qualitätseinfluss hängt von der Integration in die 3D-Szene ab. Ist eine räumliche Abgrenzung klar erkennbar (z. B. die Figur steht auf einem Fußboden; siehe Abbildung 5.3), sind bereits kleine Fehler störend. Ist diese Abgrenzung nicht deutlich, sind kleine Fehler für den Betrachter nicht erkennbar. Daher wird für die Berücksichtigung beider Fälle eine sehr kleine Schwelle gewählt. Als Resultat von Tests (siehe auch Tabelle 4.3) wird diese Schwelle zunächst mit 0,1 angesetzt. Auf Grund der Eindimensionalität eines Translationsfehlers im Bild wird eine lineare Abhängigkeit angesetzt.

³Zum Beispiel ist das Ergebnis beim Block-Matching bei starken Texturstörungen ungenau, während die Konturtransformation bei starken Skalierungen/Verzerrungen z. T. ein nicht repräsentatives Ergebnis liefert.

2. Die zeitliche Abweichung $ws_{trans,\chi}$ repräsentiert ein „Zittern“ oder ein „Rucken“ des Objektes. Dieses stört unabhängig von einer räumlichen Begrenzung, sodass lediglich die methodenbedingten Ungenauigkeiten mit einem unteren Schwellwert von 0,01 ausgeschlossen werden.
3. Kurzzeitige Störungen $wa_{trans,\chi}$ repräsentieren ein „Springen“ des Objektes. Da hier keine methodenbedingten Ungenauigkeiten auftreten, wird keine untere Schwelle angesetzt.

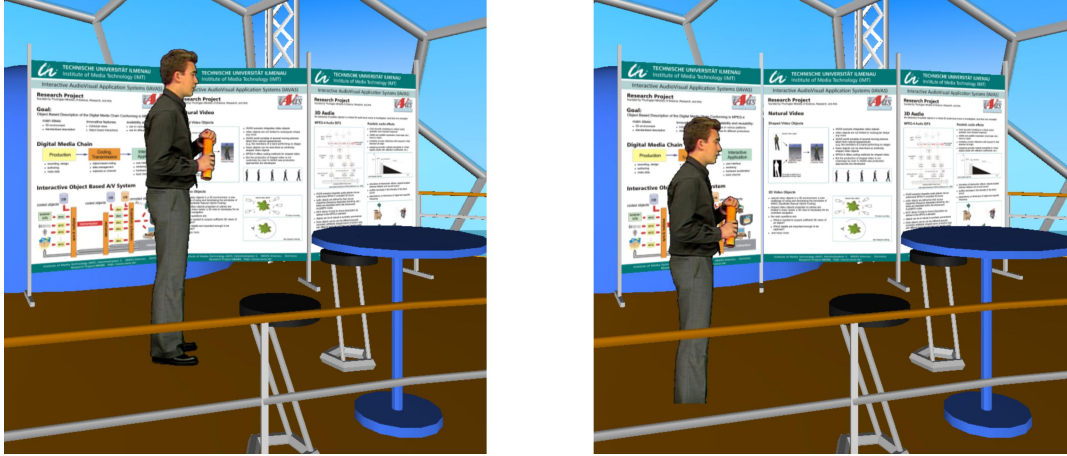


Abbildung 5.3: Sichtbarkeit eines Translationsfehlers bei räumlicher Begrenzung

Des Weiteren ist der Einfluss durch die Navigation zu berücksichtigen. Bei einer Navigation auf oder um das Objekt nav_{um} , nav_{obj} ist eine erhöhte Aufmerksamkeit zu unterstellen. Als Bezugsgeschwindigkeit wird die durchschnittliche Geschwindigkeit von Menschen beim Gehen von $v = nav_r = 6 \frac{\text{km}}{\text{h}} = 1,67 \frac{\text{m}}{\text{s}}$ genutzt. Wenn nav_{um} bzw. nav_{obj} deutlich größer als diese Bezugsgeschwindigkeit wird, ist davon auszugehen, dass die Aufmerksamkeit sinkt. Eine erhöhte Aufmerksamkeit ist selbstverständlich auch vorhanden, wenn das Objekt von einem festen Standpunkt aus betrachtet wird, d. h. $nav_{still} = 1$. Die Steigerung der Aufmerksamkeit ist jedoch von weiteren Aspekten abhängig. So wird u. a. die Wahrnehmung von Störungen gesenkt, wenn sich das Objekt bewegt. Um eine weitere Fallunterscheidung zu vermeiden, wird der Faktor der maximalen Aufmerksamkeitssteigerung mit 2 angesetzt. Bei einer beliebigen Navigation (nur $nav_v > 0$) wird nicht von einer Aufmerksamkeitssteigerung ausgegangen. Damit ergibt sich folgender Wichtungsfaktor c_{nav1} auf Grund der jeweiligen Navigation zu:

$$\begin{aligned}
 c_{nav1} &= 1 + \frac{1}{(nav_{um}/nav_r)^2 + 1} & , \text{ für } nav_{um} > 0 \\
 c_{nav1} &= 1 + \frac{1}{(nav_{obj}/nav_r)^2 + 1} & , \text{ für } nav_{obj} > 0 \\
 c_{nav1} &= 1 + nav_{still} & , \text{ für } nav_{still} > 0
 \end{aligned} \tag{5.2}$$

Die Herleitung führt somit zu folgenden drei Qualitätsparametern:

$$\begin{aligned}
 p_{position,m} &= c_{nav1} \cdot s_{0,1} \cdot \overline{wm}_{trans,\chi} \\
 p_{position,s} &= c_{nav1} \cdot s_{0,01} \cdot \overline{ws}_{trans,\chi} \\
 p_{position,a} &= c_{nav1} \cdot \overline{wa}_{trans,\chi}
 \end{aligned} \tag{5.3}$$

Größenänderungen

Im Abschnitt 4.6 wurde das Merkmalsmaß $w_{skal,kontur}$ zur Ermittlung von Skalierungsfehlern vorgestellt. Dieses Maß berücksichtigt bereits die Zunahme der Störung für Skalierungsfaktoren > 1 und $\rightarrow 0$. Die Aspekte der dynamischen Maße ähneln denen von Translationsfehlern.

1. Das zeitliche Mittel $w_{m_{skal,kontur}}$ entspricht einem permanenten Größenfehler. Sein Qualitätseinfluss hängt ebenfalls von dem Vorhandensein einer klaren räumlichen Abgrenzung ab. Daher wird wiederum eine sehr kleine Schwelle mit 0,1 angesetzt.
2. Die zeitliche Abweichung $w_{s_{skal,kontur}}$ repräsentiert ein „Atmen“ des Objektes. Dieses stört unabhängig von einer räumlichen Begrenzung, sodass lediglich die methodenbedingten Ungenauigkeiten mit einem unteren Schwellwert von 0,01 ausgeschlossen werden.
3. Kurzzeitige Störungen $w_{a_{skal,kontur}}$ repräsentieren ein „Pulsieren“ des Objektes. Da hier keine methodenbedingten Ungenauigkeiten auftreten, wird keine untere Schwelle angesetzt.

Auch der Einfluss durch die Navigation wirkt sich auf die Aufmerksamkeit in gleicher Weise wie bei Translationsfehlern aus. Daher wird wiederum der Faktor c_{nav1} genutzt.

$$\begin{aligned}
 p_{groesse,m} &= c_{nav1} \cdot s_{0,1} \cdot w_{m_{skal,kontur}} \\
 p_{groesse,s} &= c_{nav1} \cdot s_{0,01} \cdot w_{s_{skal,kontur}} \\
 p_{groesse,a} &= c_{nav1} \cdot w_{a_{skal,kontur}}
 \end{aligned} \tag{5.4}$$

Konturveränderungen

Das Merkmalsmaß $w_{silhouette,DFT}$ repräsentiert skalierungsunabhängige Störungen der Kontur der Silhouette. Dazu gehören u. a. Blickwinkelfehler und Verzerrungen beim Warping. Ein hohes Merkmalsmaß kann z. B. bei einem Blickwinkelfehler vorliegen. Die Ansicht kann dennoch beim Betrachter einen positiven Eindruck bewirken. Daher ist festzustellen, welche Qualitätsminderung durch einen Blickwinkelfehler verursacht wird. Es steht jedoch auch die Frage, inwieweit ein alleiniger Blickwinkelfehler (ohne weitere Störungen) als Fehler einer 3DVO-Generierung auftritt⁴. Bei den vorgestellten 3DVO-Verfahren würde ein alleiniger Blickwinkelfehler im Allgemeinen nur dann auftreten, wenn beispielsweise die Koordinaten einer Aufnahmekamera oder des Navigationspfades falsch zugeordnet wurden. Solch offensichtliche Fehler sollen nicht Gegenstand dieser Qualitätsbewertung sein. Man kann also davon ausgehen, dass ein Blickwinkelfehler gemeinsam mit weiteren (deutlichen) Synthesefehlern auftritt. Daher ist eine besondere Behandlung hoher Werte für $w_{silhouette,DFT}$ nicht erforderlich.

Dieses Merkmalsmaß bewertet Störungen aller Spektralanteile der Konturfunktion. Die Bewertung der Qualitätsminderung durch dieses Merkmalsmaß wird erschwert, da die Störungen völlig unterschiedlichen Charakter haben können⁵. Auf Grund der bisherigen Erfahrungen (siehe Tabelle 4.8) wird zunächst ein linearer Einfluss angesetzt.

Es werden zwei Aspekte der dynamischen Maße unterschieden:

⁴Die alleinigen Blickwinkelfehler in Abschnitt 4.6 wurden simuliert.

⁵z. B. eine unregelmäßige leichte Abweichung entlang der Kontur bei modellbasierten 3DVO-Verfahren oder lanzenförmige Warping-Fehler

1. Da sich die Kontur eines 3D-Videoobjektes ständig ändern kann, ist die zeitliche Änderung der Abweichungen $ws_{silhouette,DFT}$ kein signifikantes Maß. Daher wird das zeitliche Mittel $wm_{silhouette,DFT}$ genutzt. Für übliche Auflösungen (z. B. $d_{ref} = 576$) wird eine untere Schwelle zum Ausschluss von methodenbedingten Abweichungen⁶ mit 10^2 verwendet.
2. Kurzzeitige, besonders große Konturveränderungen lenken die Aufmerksamkeit des Betrachters auf die Störung des Objektes und werden daher mittels $wa_{silhouette,DFT}$ erfasst. Da hier keine methodenbedingten Abweichungen auftreten, wird keine untere Schwelle angesetzt.

Auch der Einfluss einer eventuell vorhandenen Navigation wirkt auf die Aufmerksamkeit in gleicher Weise wie bei Translations- und Skalierungsfehlern aus. Daher wird wiederum der Faktor c_{nav1} genutzt.

$$\begin{aligned} p_{kontur,m} &= c_{nav1} \cdot s_{100} \cdot wm_{silhouette,DFT} \\ p_{kontur,a} &= c_{nav1} \cdot wa_{silhouette,DFT} \end{aligned} \quad (5.5)$$

Eine weitere Klasse an Merkmalsmaßen erfasst Topologieänderungen, insbesondere die Veränderung von Silhouettenöffnungen: $w_{\#silhouette}$ und $w_{\Delta silhouette}$. Da noch keine ausführlichen subjektiven Tests zur Bestimmung der Qualitätsbeeinflussung vorliegen können, werden die Untersuchungsergebnisse von vergleichbaren Segmentierungsartefakten genutzt. In [Gel04] wurde der Einfluss von Topologieänderungen der Segmentierung auf die Qualität untersucht. Insbesondere wurden Löcher und hinzugefügte Bereiche als Störungstypen analysiert. Mittels subjektiver Testverfahren nach [ITU02] und [ITU99] wurde der mittlere Störwert (*MAV, mean annoyance value*) bestimmt. Die Analyse der Testdaten führte zu einem logarithmischen Zusammenhang⁷. Daher wird für die Bildung der Qualitätsparameter ebenfalls von einem logarithmischen Zusammenhang ausgegangen.

Es müssen zwei dynamische Aspekte des Merkmalsmaßes $w_{\#silhouette}$ (basierend auf der Anzahl der Öffnungen) berücksichtigt werden:

1. Das zeitliche Mittel $wm_{\#silhouette}$ entspricht dem Gesamtausmaß solcher Topologiefehler. Da in den meisten Fällen Löcher bereits in Bildpunktgröße sichtbar sind, ist keine Schwellenwertbildung erforderlich.
2. Die zeitliche Abweichung $ws_{\#silhouette}$ repräsentiert Veränderungen von Topologieänderungen, z. B. das wiederholte Entstehen und Verschwinden von Löchern. Da solche zeitlichen Veränderungen die Aufmerksamkeit auf sich ziehen, muss dieses Verhalten ebenfalls erfasst werden.

Die Gewichtung auf Grund der Navigation erfolgt wieder mit dem Faktor c_{nav1} .

$$\begin{aligned} p_{topologie\#,m} &= c_{nav1} \cdot \lg(wm_{\#silhouette} + 1) \\ p_{topologie\#,s} &= c_{nav1} \cdot \lg(ws_{\#silhouette} + 1) \end{aligned} \quad (5.6)$$

Die Berücksichtigung der dynamischen Aspekte des Merkmalsmaßes $w_{\Delta silhouette}$ (basierend auf den Größenabweichungen der Öffnungen) erfolgt unter folgenden Gesichtspunkten:

⁶z. B. auf Grund des Schwellenwertes für Alpha zur Bestimmung der Kontur; Bestimmung der Konturfunktion

⁷Der Zusammenhang wird mit $y = a + b \cdot \lg(x + c)$ angegeben, mit z. B. $a = -15,75$; $b = 8,87$; $c = 5,95$.

1. Das zeitliche Mittel $wm_{\Delta silhouette}$ entspricht den mittleren Größenfehlern der Öffnungen. Da bei kleinen Öffnungen bereits kleine Fehler sichtbar sind, ist keine Schwellenwertbildung erforderlich.
2. Die zeitliche Abweichung $ws_{\Delta silhouette}$ repräsentiert zeitliche Größenveränderungen der Öffnungen, z. B. temporäre Vergrößerungen während der Navigation (jeweils in den Positionen zwischen den Aufnahmekameras). Da solche zeitlichen Veränderungen die Aufmerksamkeit auf sich ziehen, muss dieses Verhalten ebenfalls erfasst werden.

$$\begin{aligned} p_{topologie\Delta,m} &= c_{nav1} \cdot \lg(wm_{\Delta silhouette} + 1) \\ p_{topologie\Delta,s} &= c_{nav1} \cdot \lg(ws_{\Delta silhouette} + 1) \end{aligned} \quad (5.7)$$

Lokale Veränderungen statistischer Größen

Mehrere Störungen bewirken lokale Änderungen des örtlichen Informationsgehaltes und des Histogramms. Hierzu gehören Texturierungsfehler, Okklusionen, Warping-Fehler, Blickwinkelfehler u. a. (siehe Abschnitt 4.6). Zur Erfassung dieses Einflusses dienen die Merkmalsmaße $w_{blockSI}$ und $w_{blockHI}$. Abhängig von den Eigenschaften des Bildes führt die Detektion durch das gradientenbasierte Maß $w_{blockSI}$ oder durch das farbverteilungsbasierte Maß $w_{blockHI}$ zu einem besseren Ergebnis. Die Erfahrungen zeigen jedoch, dass für viele Objekte beide Maße ähnlich geeignet sind (siehe u. a. Tabelle 4.7). Es erfolgt daher eine Wichtung wie für $\bar{w}_{trans,\chi}$:

$$\bar{w}_{blockSI,HI} = \frac{w_{blockSI} + w_{blockHI}}{2 \cdot ((w_{blockSI} - w_{blockHI})^2 + 1)} \quad (5.8)$$

Für die Erfassung des Zusammenhanges dieser Merkmalsmaße mit der Qualität gelten ähnliche Voraussetzungen wie für die Bildung von $p_{kontur,m}$ und $p_{kontur,a}$ ⁸. Die Qualitätsparameter werden daher in gleicher Weise gebildet:

$$\begin{aligned} p_{lokal,m} &= c_{nav1} \cdot s_{0,1} \cdot \bar{w}_{blockSI,HI} \\ p_{lokal,a} &= c_{nav1} \cdot \bar{w}_{blockSI,HI} \end{aligned} \quad (5.9)$$

Globale Veränderungen statistischer Größen

Störungen, welche die statistischen Größen der gesamten Ansicht beeinflussen, können mit w_{SI} und w_{HI} erfasst werden. Typische Beispiele hierfür sind großflächige Artefakte auf Grund einer zu hohen Texturdatenkompression oder Abweichungen infolge einer signalseitig dejustierten Kamera. Selbstverständlich können auch größere Perspektivfehler zu einem Anstieg dieser Maße führen. Da die Beeinflussung der Qualität mit der von den lokalen statistischen Merkmalsmaßen vergleichbar ist, werden die globalen statistischen

⁸Die Bildung der Qualitätsparameter und deren RCA in [Wol02] zeigen, dass eine detailliertere Anpassung der statistischen Maße mittels Bewertungsfunktionen erforderlich ist. Es ist jedoch zu berücksichtigen, dass die Störungstypen bei 3D-Videoobjekten weitaus vielfältiger sind und daher nur begrenzt durch Bewertungsfunktionen unterscheidbar sind.

Qualitätsparameter genauso gebildet:

$$\begin{aligned}\bar{w}_{SI,HI} &= \frac{w_{SI} + w_{HI}}{2 \cdot ((w_{SI} - w_{HI})^2 + 1)} \\ p_{global,m} &= c_{nav1} \cdot s_{0,1} \cdot \bar{w}_{SI,HI} \\ p_{global,a} &= c_{nav1} \cdot \bar{w}_{SI,HI}\end{aligned}\tag{5.10}$$

5.2 Bildung eines 3D-Videoobjektqualitätsmaßes

5.2.1 Zusammenführung von Qualitätsparametern

Problemstellung

Im vorherigen Abschnitt wurde die Bildung einer Auswahl von relevanten Qualitätsparametern p_ψ vorgenommen. Im nächsten Schritt werden die Qualitätsparameter zu einem 3D-Videoobjektqualitätsmaß 3DVQM zusammengefasst (siehe Abbildung 5.1). Dieser Schritt ist allerdings nicht für alle Anwendungen erforderlich. Wenn beispielsweise die Abhängigkeit einer einzelnen Störgröße untersucht werden soll (z. B. durch eine hohe Kompression) kann dies mittels einzelner oder einer Auswahl an Qualitätsparametern erfolgen. Die Zusammenführung zu einem einzigen Qualitätsmaß ist jedoch immer erforderlich, wenn mehrere Störungen wirksam sind oder ein Gesamtprädikat für das Videoobjekt gewünscht ist.

Für die Zusammenführung von Qualitätsparametern muss der unterschiedliche Einfluss der einzelnen Merkmalsmaße auf die subjektive Bewertung quantitativ erfasst werden. Hierzu sind jedoch folgende Aspekte zu beachten:

- Die subjektive Bewertung der einzelnen Qualitätsmerkmale bzw. Störungen ist individuell verschieden. Erste Erfahrungen bei der Nutzung von 3D-Videoobjekten zeigen, dass die Nutzer ihr Augenmerk auf verschiedene Aspekte richten⁹. Beispielsweise wurde von einigen Nutzern auf die bei der Navigation sich ändernden Beleuchtungsverhältnisse geachtet, während andere Nutzer durch Okklusionen abgelenkt wurden.
- Die subjektive Bewertung der einzelnen Qualitätsmerkmale bzw. Störungen ist abhängig von der Nutzungssituation, der Applikation und den Endgeräten. 3D-Videoobjekte stellen künftig einen neuen Medienobjekttyp für verschiedene Applikationen (z. B. Free Viewpoint Video *FVV*, Free Viewpoint TV *FTV*, 3D-TV, Web-Applikationen) dar. Hierzu werden unterschiedliche Endgeräte (z. B. PC, Spielekonsole, Fernsehgerät) genutzt. Diese Faktoren beeinflussen die Wahrnehmung und damit die Bewertung des Betrachters¹⁰.
- Die subjektive Bewertung der einzelnen Qualitätsmerkmale bzw. Störungen ist in einem hohen Maße vom verwendeten Videoobjekt abhängig¹¹. So haben Störungen

⁹Dieses Verhalten ist natürlich auch der Tatsache geschuldet, dass bei den Nutzern eine Erwartungshaltung aufgebaut wurde, bevor sie die für sie noch kaum bekannten 3D-Videoobjekte betrachtet haben.

¹⁰In [Wol02] werden daher für die unterschiedlichen Anwendungen von konventionellem Video (TV, Videokonferenz etc.) unterschiedliche Auswahlen an Qualitätsparametern mit unterschiedlichen Gewichtungsfaktoren zusammengeführt.

¹¹Der Einfluss des Inhaltes auf die Bewertung ist insbesondere für die Auswahl von Testmaterial signifikant (z. B. [ANSI95]). Er wurde u. a. auch für stereoskopische Aufnahmen in [Kay96] nachgewiesen.

einen unterschiedlichen Einfluss je nachdem, ob das Videoobjekt beispielsweise einen agierenden oder einen passiven Menschen abbildet.

- Die subjektive Bewertung der einzelnen Qualitätsmerkmale bzw. Störungen ist auch von der Umgebung des Videoobjektes abhängig. Dieser Aspekt muss jedoch für die alleinige Bewertung von 3D-Videoobjekten ausgeschlossen werden.

Methoden für die Bildung eines einzelnen Maßes

In vielen Arbeiten zur Qualitätsbewertung von Bewegtbild bzw. Segmentierung wird versucht, die objektive Bewertung mit einem einzigen Wert zu formulieren [Cor01], [Cor03], [Erd01], [Fec98], [Fec00], [Gel04], [Mas04], [Rav98], [Sar01], [Sar02], [Sch94], [Tra98], [Tur04], [Vil99], [VQEG03], [Win00], [Wol98], [Wol02] u. a. Die Bildung dieses einen Wertes erfolgt jedoch auf unterschiedliche Weise. Für HVS-basierte Maße ergibt sich ein einzelner Wert bereits aus dem Grundprinzip, dass man das visuelle System vollständig nachzubilden versucht [Fec98], [Sar01], [Sar02], [Gu99]. Für die Bildung eines merkmalsbasierten Maßes muss der Einfluss der einzelnen Störmaße/Qualitätsparameter mittels multipler Regression erfasst werden [Fec00]. Verschiedene Untersuchungen (z. B. [VQEG03]) zeigen, dass diese Regression mittels subjektiver Tests und Erfahrungswerten erfolgen kann. Die Regression kann aber auch mit Hilfe von Methoden der Statistik erfolgen, bei denen die Abhängigkeit der Variablen (Störmaße/Qualitätsparameter) erfasst wird. Mit der so genannten Faktorenanalyse wird die hohe Dimensionalität des Untersuchungsraumes¹² durch wenige, voneinander unabhängige Variablen ersetzt. Diese Methode wurde in [Kay96] für die Beurteilung (mit einer NR-Methode) von datenkomprimierten stereoskopischen Fernsehbildern angewendet. Dort wurden neun Beurteilungsaspekte mittels Cluster-/Faktorenanalyse auf drei relevante Aspekte reduziert (Schärfe, Rauschen, Blockausfälle).

Charakter der 3DVO-Qualitätsparameter

Für die Zusammenführung der 3DVO-Qualitätsparameter ist zunächst festzustellen, dass im Vergleich zum konventionellen Bewegtbild eine größere Vielfalt an Qualitätsaspekten bzw. Störungen zu erfassen ist¹³. Aus diesem Grund wird die Anpassung eines einzelnen Maßes an die subjektive Bewertung tendenziell ungenauer. Beispielsweise ist es individuell verschieden, ob ein 3D-Videoobjekt mit deutlichen Konturveränderungen oder eines mit deutlichen Synthesefehlern als subjektiv schlechter bewertet wird, während das objektive Maß beide gleich bewerten würde. Die folgenden Beispiele verdeutlichen die Unterschiedlichkeit der Störungen, die zu den entworfenen Qualitätsparametern führen:

- $p_{position,a}$: oszillierende Abweichungen von der Position („Wackeln“ oder „Zittern“ des gesamten Objektes)
- $p_{kontur,a}$: Ausfransungen des Objektes; Synthesefehler entlang der Kontur (z. B. spitze Ausbuchtungen)
- $p_{topologie\#,s}$: Löcher im Objekt (Der Hintergrund ist an dieser Stelle sichtbar.)
- $p_{lokal,a}$: Okklusionen (bzw. deren Retuschen); verzerrte Objektbereiche

¹²d. h. viele wechselseitig hoch korrelierende Variablen

¹³Zudem wird auch bei der merkmalsbasierten Videoqualitätsbeurteilung versucht, die Beurteilung auf wenige Aspekte (z. B. Unschärfe, Rauschen) zu reduzieren.

- $p_{global,m}$: hohe Kompression der Texturdaten
- $p_{global,a}$: Veränderungen von Helligkeit und Kontrast des Objektes während der Navigation

Diese Aufstellung zeigt, dass ein 3D-Videoobjektqualitätsmaß nur eine begrenzte Aussage über die Vergleichbarkeit der Qualität von verschiedenen 3D-Videoobjekten geben kann.

- *Conclusio 5.5*: Die Zusammenführung von mehreren Qualitätsparametern zu einem einzigen Qualitätsmaß stellt für 3D-Videoobjekte einen größeren Kompromiss dar, da die Qualitätsaspekte und Störungen deutlich vielfältiger sind.

Die Vielfalt der in dieser Arbeit berücksichtigten Störungen ist Folge der unterschiedlichen 3DVO-Verfahren und deren teilweise frühen Entwicklungsstände. Natürlich ist davon auszugehen, dass mit der Entwicklung der 3DVO-Verfahren sich nur wenige Verfahren etablieren werden und durch deren Reife einige Störungen ausgeschlossen werden können. Anhand der aktuellen Entwicklungen (z. B. [Smo04], [Lam04], [Zie04]) lässt sich ein Prognose anstellen, auf welchen Umfang sich künftig Qualitätsparameter/Störmaße reduzieren lassen:

- $p_{kontur,a}$: Ausfransungen und fehlende Teile des Objektes
- $p_{lokal,a}$: Okklusionen und fehlerhaft texturierte Objektbereiche
- $p_{global,m}$: Kompression der Texturdaten

Diese beispielhafte Prognose verdeutlicht, dass einerseits weniger Qualitätsparameter zu berücksichtigen sind und andererseits aber auch angepasstere Qualitätsparameter entwickelt werden können.

- *Conclusio 5.6*: Mit der zunehmenden Reife von 3DVO-Verfahren können künftig einige Störungen ausgeschlossen werden¹⁴. Demzufolge werden objektive Maße künftig höher mit der subjektiven Bewertungen korrelieren können.

Selbstverständlich ist auch zu erwarten, dass durch Neu- und Weiterentwicklungen der 3DVO-Verfahren neuartige Störungen erfasst werden müssen oder auch Störungen erfasst werden müssen, die erst für hohe Qualitätsansprüche relevant werden. Für eine Erweiterung um die dafür notwendigen Parameter ist die hier vorgestellte Methodik jedoch grundsätzlich geeignet (siehe Unterabschnitt 5.1.2).

Zusammenführbarkeit der 3DVO-Qualitätsparameter

Die Zusammenführung der 3DVO-Qualitätsparameter erfordert folgende Voraussetzungen:

1. *Addierbarkeit*
2. *Kommutativität*
3. *Linearität*

¹⁴Diese Vorgehensweise ist z. B. auch für die Qualitätsbewertung von konventionellem Video üblich. Die Verfahren werden entsprechend dem HRC (*hypothetical reference circuit*) und SRC (*source sequence*) entwickelt und bewertet [VQEG03], [Ahr98], [Rav98], [Tra98].

Die Addierbarkeit der Parameter ist gegeben, wenn die zu Grunde liegenden Störungen den subjektiven Eindruck auch in ihrer Summe beeinflussen. Dieser Zusammenhang ist durch die Verifikation der merkmalsbasierten Methoden zur Videoqualitätsbeurteilung bereits nachgewiesen [VQEG03] u. a. In [Fec98] wird darauf hingewiesen, dass viele Bildqualitätsmodelle nur für den Bereich der Schwellenwahrnehmung gültig sind. Stark überschwellige Störungen führen zu Sättigungseffekten in der menschlichen Wahrnehmung, sodass die Addierbarkeit der Störmaße nicht mehr gegeben ist. Diese Fälle sind jedoch für die Qualitätsbeurteilung von konventionellem Bewegtbild jedoch kaum noch relevant. In [Kay96] wird die Addierbarkeit von Qualitätsaspekten bei der menschlichen Wahrnehmung mit der Addierbarkeit der Einzelleistungen des menschlichen Gehirnes zur Bestimmung der Intelligenz verglichen und nachgewiesen.

Weiterhin ist zu klären, ob für die in dieser Arbeit entworfenen Qualitätsparameter¹⁵ $p_{\psi,1}, p_{\psi,2}, \dots, p_{\psi,N}$ eine gleiche Verfahrensweise (d. h. eine gewichtete Summierung) anwendbar ist, wie sie für die Qualitätsparameter anderer merkmalsbasierter Maße angewendet wird. Die Gegenüberstellung in Tabelle 5.1 verdeutlicht, dass sich die

Qualitätsaspekte/-merkmale, Störmaße verschiedener merkmalsbasierter Maße	Unschärfe, Bildrauschen Blockausfälle Kantenrauschen Flimmern, Kantenflimmern Kontrastverlust, Farbfehler
Qualitätsmerkmale, Störmaße für das 3DVQM	Größenänderung Konturstörungen Ortsänderungen lokale Änderung statistischer Größen globale Änderung statistischer Größen

Tabelle 5.1: Gegenüberstellung von Qualitätsaspekten, -merkmalen und 3DVO-Qualitätsparametern

Qualitätsaspekte/-merkmale von konventionellem Video in zwei Klassen einteilen lassen. Die eine Klasse sind Maße, die einen konkreten Störungstyp messen, der technisch gegeben ist und vom Menschen direkt wahrgenommen wird. Ein typisches Beispiel hierfür sind Blockausfälle, welche seit der Einführung der Datenreduktion zu einem typischen Störungstyp wurden. Die andere Klasse sind Maße, welche die Auswirkungen von meist mehreren ähnlichen Störungen messen. Ein typisches Beispiel ist das Bildrauschen. Es kann mehrere Ursachen haben und wird unabhängig von der Ursache als Störung wahrgenommen¹⁶. Die in dieser Arbeit vorgestellten Qualitätsparameter repräsentieren in gleicher Weise Qualitätsaspekte bzw. Störungen beider Klassen. Die Parametergruppen $p_{position}$, $p_{groesse}$, p_{kontur} und $p_{topologie}$ erfassen unmittelbar Störungen, die vom Betrachter direkt wahrgenommen werden. Die Parametergruppen p_{lokal} und p_{global} erfassen die statistisch messbaren Auswirkungen verschiedener Fehler (Okklusionen, Synthesefehler, Kompressionsartefakte etc.), welche durch den Betrachter mittelbar wahrgenommen werden¹⁷. Daher kann für die 3DVO-Qualitätsparameter von einer Addierbarkeit ausgegangen werden.

¹⁵Dies umfasst die hier exemplarisch entworfenen Parameter sowie (künftige) Parameter, welche nach demselben Prinzip entwickelt worden sind.

¹⁶Die Charakteristik des Rauschens (z. B. weiß) hat jedoch Einfluss auf die Wahrnehmung.

¹⁷Der Einfluss auf die Wahrnehmung wird nur in einem begrenzten Maße quantitativ mit den statistischen Maßen erfassbar sein. Bei einer künftigen Einschränkung der Störungen und einer höheren

Die Kommutativität der Parameter ist gegeben, wenn die Parameter $p_{\psi,1}, p_{\psi,2}, \dots, p_{\psi,n}$ für das Qualitätsmaß austauschbar sind. In diesem Fall entspricht ein Qualitätsmaß einer bestimmten Größe immer der gleichen subjektiven Bewertung unabhängig von der Verteilung der Parameter. Dieses Problem wurde bereits diskutiert und die sich ergebenden Einschränkungen erläutert.

Eine Linearität der Parameter ist gegeben, wenn die Parameter zueinander einen linearen Einfluss auf die Qualitätsbewertung widerspiegeln¹⁸. Hierfür wurde bei der Modellierung der Qualitätsparameter auf die Einhaltung des Weber-Fechner-Gesetzes geachtet (z. B. Größenänderungen). Für Parameter, deren Einfluss auf die Wahrnehmung nicht mit der Intensität einer physikalischen Größe vergleichbar ist, wurden Untersuchungen durchgeführt bzw. auf andere Untersuchungen zurückgegriffen (z. B. Topologieänderungen). Dennoch ist davon auszugehen, dass die Linearität nicht für alle Parameter in ihrem gesamten Wertebereich gegeben ist. Insbesondere sind Nichtlinearitäten durch den Einfluss unterschiedlicher Störungen, die zu einem einzigen (statistischen) Qualitätsparameter führen, zu erwarten.

Gewichtete Summierung

Es wurde gezeigt, dass sich das 3D-Videoobjektqualitätsmaß als gewichtete Summe der Qualitätsparameter p_{ψ} bestimmen lässt:

$$VQM = \sum_{i=1}^N \beta_i \cdot p_{\psi,i} \quad (5.11)$$

Die Genauigkeit des Videoobjektqualitätsmaßes hängt in einem hohen Maße von der Wahl der Koeffizienten β_i ab. Wenn die Qualitätsparameter $p_{\psi,i}$ erweitert oder modifiziert werden sollen, sind selbstverständlich auch alle Koeffizienten anzupassen.

5.2.2 Methodik zur Bestimmung der Gewichtungskoeffizienten

Voraussetzungen

Für die Bestimmung der Koeffizienten von merkmalsbasierten Maßen wurde in anderen Arbeiten die Cluster-/Faktorenanalyse erfolgreich angewendet (z. B. [Kay96]). Hierfür sind folgende Voraussetzungen nötig:

1. statistische Untersuchungen zu den Wahrnehmungsaspekten zur Bestimmung der Dimensionalität
2. mathematische Modelle zur Erfassung/Messung der relevanten Aspekte

Beide Voraussetzungen sind für das 3D-Videoobjektqualitätsmaß nur teilweise erfüllt. Die in 1.) genannten statistischen Untersuchungen erfordern ausreichend viele 3D-Videoobjekte, welche die relevanten Qualitätsaspekte in einer Abstufung aufweisen. Da die Entwicklung von Verfahren für natürliche 3D-Videoobjekte sich noch am Anfang befindet, liegt die erforderliche Vielfalt noch nicht vor. Auch die in dieser Arbeit vorgestellten

Angepasstheit der statistischen Maße kann die Übereinstimmung mit der Wahrnehmung verbessert werden.

¹⁸Das Qualitätsmaß als gewichtete Summe der Parameter wird über eine nichtlineare Transformationen an den MOS bzw. DMOS angepasst.

Qualitätsparameter erfüllen die Bedingung 2.) nur teilweise, da einige Aspekte sich nur in Kombination oder auch teilweise in mehreren Qualitätsparametern wiederfinden¹⁹.

Aus den genannten Gründen wird für die Bestimmung der Koeffizienten des 3DVQM eine sukzessive Approximation und Regression auf Basis erweiterter subjektiver Bewertungsreihen vorgeschlagen. Mit dieser Methode wird der Forderung nach einer künftigen Anpassbarkeit des 3D-Videoobjektqualitätsmaßes an spätere Entwicklungen und Anforderungen genügt. In Abbildung 5.4 ist die vorgeschlagene Methode zur sukzessiven Ap-

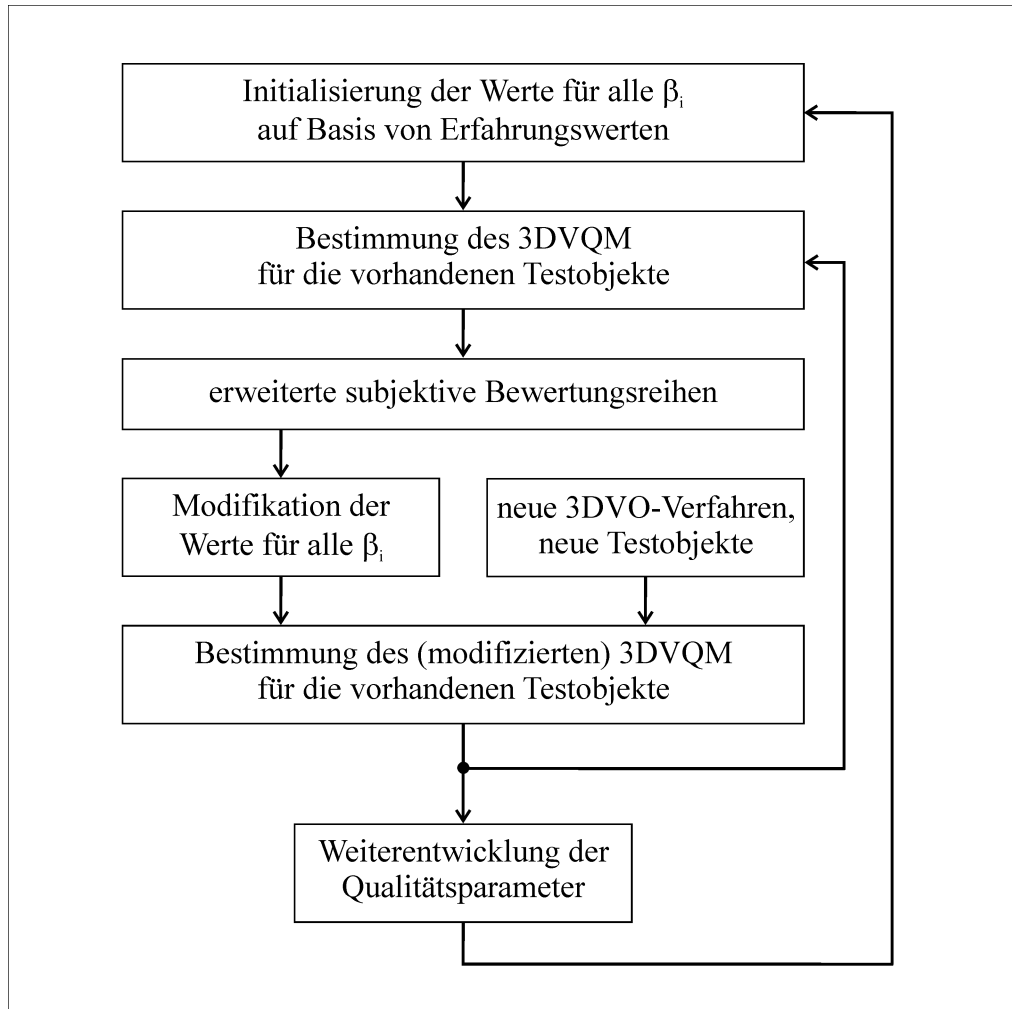


Abbildung 5.4: Bestimmung der Koeffizienten für das 3DVQM

proximation auf Basis erweiterter subjektiver Bewertungsreihen dargestellt. Die Methode berücksichtigt sowohl die sukzessive Anpassung der Koeffizienten β_i als auch die Ergänzung und Weiterentwicklung der Qualitätsparameter p_ψ .

Erweiterte subjektive Bewertung

Die subjektive Bewertung muss für die Anwendung der oben genannten Methode um Bewertungsaspekte ergänzt werden, mittels derer ein Rückschluss auf die Wichtung der zur Verfügung stehenden Qualitätsparameter möglich ist. Im ersten Schritt ist jedoch die Methode zur subjektiven Bewertung der Qualität festzulegen. Hierfür sind die vielfältigen

¹⁹Beispiel: Wenn ein Objekt an einer Stelle eine Auswuchtung aufweist, ändern sich sowohl die Kontur als auch die lokalen statistischen Maße.

Erfahrungen, die für die subjektive Bewertung von konventionellem Bewegtbild vorliegen, zu nutzen.

Da es sich bei dem zu entwickelnden 3D-Videoobjektqualitätsmaß um ein vergleichendes Maß²⁰ handelt, ist eine Double-Stimulus-Methode zu verwenden. Für erste Untersuchungen sind Methoden sinnvoll, welche zu einer Gesamtbewertung der Sequenz führen²¹. Bei der DSCQS-Methode weiß der Betrachter nicht, welche der beiden Sequenzen die Referenz ist. Dadurch können kleinere Unterschiede erfasst werden. Diese Vorgehensweise ist für die Bewertung von 3D-Videoobjekten noch nicht geeignet, da die Betrachter ihre Konzentration auf den neuartigen Medienobjekttyp und dessen Störungen richten sollen. Aus diesen Gründen wird die DSIS-Methode zur Präsentation der Videoobjekte genutzt. Daher wird zunächst immer die Referenz und anschließend das gestörte Objekt gezeigt. Der für die DSIS-Methode festgelegte zeitliche Ablauf (Referenz 10 s, Pause 3 s, zu bewertende Sequenz 10 s) sollte im ersten Schritt als Orientierung dienen, solange keine genaueren Kenntnisse über die Wahrnehmungsabläufe von 3D-Videoobjekten vorliegen.

Die Präsentation von 3D-Videoobjekten kann im Gegensatz zu konventionellem Video sowohl passiv als auch interaktiv erfolgen. Da ein Vergleich mit der Referenz erfolgen soll, ist zu gewährleisten, dass beide Videoobjekte durch dieselbe Blickpunktwahl entstanden sind. So wäre eine Präsentation denkbar, bei welcher der Betrachter/Nutzer²² die Referenz durch eine eigene Navigation bestimmt und das gestörte Videoobjekt für die gewählte Navigation präsentiert wird. Abgesehen von dem enormen Aufwand hat diese Methode den Nachteil, dass der Nutzer bei der Referenz aktiv und beim gestörten Objekt passiv ist. Damit wäre die Vergleichbarkeit fraglich. Aus diesen Gründen werden die 3D-Videoobjekte passiv für eine typische Nutzung (d. h. Navigation) präsentiert²³.

Die Bewertung bei der gewählten DSIS-Methode erfolgt nach einer fünfstufigen Skala (siehe Tabelle 2.2). Diese setzt indirekt voraus, dass die Referenz als ungestört angesehen werden kann. Dies kann zum jetzigen Entwicklungsstand von 3D-Videoobjekten nicht vorausgesetzt werden. Daher wird die DSCQS-Skala (siehe Abbildung 2.9) verwendet. Damit ist eine differentielle Bewertung möglich, sodass die subjektive Bewertung der Referenz einbezogen wird. Auch erlaubt die DSCQS-Methode eine feinere Abstufung. Im Bewertungsformular werden deutschsprachige Begriffe gemäß [Tei90] für die Qualitätsbewertung verwendet.

Die Erweiterung der subjektiven Bewertung ist notwendig, um die für die subjektive Bewertung relevanten Aspekte vom Betrachter zu erfahren. Mit diesen Informationen wird die sukzessive Anpassung der Koeffizienten β_i vorgenommen. Hierfür gibt es im Bewertungsformular Skalen für die einzelnen Aspekte. In Anbetracht des Umfangs der Bewertung einschließlich derer Einzelaspekte kann nur eine grobe Skala zur Bewertung des Ausmaßes der Störungen zur Anwendung kommen. Es wird eine dreistufige Skala verwendet, welche in „stark störend“, „störend“ und „nicht störend“ unterteilt ist. Des Weiteren muss der Einfluss der zeitlichen Charakteristik der Störungen auf die Wahrnehmung erfasst werden. Im Unterabschnitt 4.7.1 wurden drei zeitliche Eigenschaften untersucht und in die Qualitätsbewertung einbezogen (beschrieben durch: arithmetisches Mittel, Stan-

²⁰Vergleich mit einer Referenz (siehe Abschnitt 4.3)

²¹Für künftige Untersuchungen sind auch kontinuierliche Bewertungsmethoden (*continuous evaluation*, z. B. SDSCE) zu erproben. Dies setzt jedoch Erfahrungen in der Bewertung sowie Erfahrungen der Testpersonen voraus.

²²Da der Betrachter von 3D-Videoobjekten in diesem Falle agiert, wird er auch als Nutzer bezeichnet.

²³Bei der Entwicklung der Methodik für die objektive Bewertung wurde ebenfalls von einer typischen Nutzung ausgegangen.

dardabweichung, kurzzeitige Maxima). Um die Anzahl der Kriterien der subjektiven Bewertung gering zu halten, werden die zeitlichen Eigenschaften lediglich in „permanent“ und „kurzzeitig“ unterschieden. Dies ist möglich, da die Charakterisierung eines permanenten Fehlers mit Hilfe des arithmetischen Mittels oder der Standardabweichung zumeist offensichtlich ist. In Abbildung 5.5 ist ein Beispiel für das Bewertungsformular dargestellt.

- *Conclusio 5.7:* Die erweiterte subjektive Bewertung setzt Kenntnisse über die Störungstypen von 3D-Videoobjekten beim Betrachter voraus. Diese Art der Expertenbewertung ist in dieser frühen Phase der Qualitätsbewertung zulässig²⁴.

Videoobjekt 1					
Qualitätsbewertung:	Ref.	Test			
			ausgezeichnet		
	+	+	gut		
	+	+	ausreichend		
	+	+	schlecht		
			sehr schlecht		
Bewertung der Störungen:					
	stark störend	störend	nicht störend	permanent	kurz- zeitig
fehlerhaft erscheinende Objektbereiche	----- ----- -----			----- -----	
Konturfehler	----- ----- -----			----- -----	
Positionsfehler / Unruhe	----- ----- -----			----- -----	
Unschärfe des gesamten Objektes	----- ----- -----			----- -----	

Abbildung 5.5: Formular für die erweiterte subjektive Bewertung

Die Auswahl der im Bewertungsformular genannten Störungen hat entscheidenden Einfluss auf die Verwendbarkeit der Ergebnisse. Hierfür lassen sich die folgenden Anforderungen erkennen:

- Die Anzahl der Störungstypen darf nicht zu groß gewählt werden, da dies den Bewerter überfordern könnte. Daher ist auf Störungstypen, welche für das zu untersuchende 3D-Videoobjekt offensichtlich irrelevant sind, zu verzichten.

²⁴In [ITU99] wird auf die Eignung von Bewertungen durch kleine Expertengruppen (4...8) in frühen Phasen der Systementwicklung hingewiesen.

- Es ist ein Optimum zwischen der Anzahl und der Genauigkeit der aufgelisteten Störungen zu finden. Je detaillierter die Störungstypen genannt sind, desto genauer kann der Bezug zu den Qualitätsparametern und dessen Koeffizienten β_i hergestellt werden.
- Die sprachliche Formulierung der Störungstypen muss so gewählt werden, dass die Betrachter unmissverständlich wissen, welche Art von Störungen gemeint ist. Dies ist mitunter schwierig, da sich für die speziellen Störungen von 3D-Videoobjekten noch keine prägnanten Begriffe etabliert haben.

Initialisierung der Koeffizienten

Bei der Initialisierung müssen Werte für die Koeffizienten β_i auf Basis der erworbenen Erfahrungen festgelegt werden. Als Kriterium hierfür steht lediglich die Kommutativität, d. h. mittels der Wichtung muss jeder Qualitätsparameter den gleichen Einfluss innerhalb des 3DVQM bekommen. Als Anhaltspunkt eignet sich die Schwelle, ab der man eine spürbare Beeinträchtigung wahrnimmt und diese auf z. B. 1 skaliert.

Modifikation der Koeffizienten

Mit Hilfe des erweiterten subjektiven Testes kann ein Rückschluss darauf gezogen werden, welche Qualitätsaspekte für die subjektive Bewertung des jeweiligen Testobjektes entscheidend gewesen sind. Aus den Bewertungen mehrerer Betrachter lässt sich eine Wichtung der Aspekte für die subjektive Wahrnehmung abschätzen. Diese kann wiederum mit den Anteilen der Qualitätsparameter²⁵ $\tilde{\alpha}_i$ verglichen werden. Selbstverständlich ist eine einzige Bewertungsreihe nicht ausreichend, die Gewichtungskoeffizienten zu bestimmen, da bei anderen 3D-Videoobjekten sich die vielfältigen Phänomene der menschlichen Wahrnehmung anders auswirken können. Bei der hier verwendeten sukzessiven Approximation werden die Wichtungskoeffizienten mit jedem Bewertungszyklus korrigiert.

- *Conclusio 5.8:* Die sukzessive Approximation der Gewichtungskoeffizienten erlaubt eine schrittweise Bildung des 3D-Videoobjektqualitätsmaßes entsprechend der Entwicklung von 3DVO-Verfahren und der zur Verfügung stehenden Testobjekte.

Für die Modifikation der Koeffizienten müssen zunächst die Ergebnisse der erweiterten subjektiven Bewertungsreihen ausgewertet werden. In einem ersten Schritt ist aus den Qualitätsbewertungen der DMOS (*differential mean opinion score*) zu bestimmen. Aus der Markierung in der Bewertungsskala wird ein ganzzahliger Wert im Bereich $0 \dots 100$ bestimmt. Die Differenz der Werte für Referenz und Messobjekt ergibt die differentielle Bewertung u_{roh} . Die Einzelbewertungen u_{roh} werden zunächst skaliert²⁶ (nach [VQEG03]):

$$u = \frac{u_{roh} - u_{min}}{u_{max} - u_{min}} \quad (5.12)$$

Die Werte u_{max} und u_{min} sind jeweils die höchste und geringste differentielle Bewertung eines Betrachters. Durch diese Skalierung werden die Bewertungen der einzelnen Betrachter auf das Intervall $0 \dots 1$ skaliert. Die sonst üblichen Untersuchungen, ob Testpersonen ausgeschlossen werden, können bei dieser kleinen Expertentestreihe noch nicht durchgeführt

²⁵Produkt aus Wichtungskoeffizient und Qualitätsparameter im Verhältnis zum Qualitätsmaß $\tilde{\alpha}_i = \frac{\beta_i \cdot p_{\psi,i}}{VQM}$

²⁶Die Korrektur einer absoluten Abweichung ist bei einer differentiellen Bewertung nicht notwendig.

werden. Der DMOS für ein Objekt ergibt sich als arithmetisches Mittel der Bewertungen der Betrachter.

$$DMOS = \frac{1}{N} \sum_{i=1}^N u_i \quad (5.13)$$

Im zweiten Schritt sind die subjektiven Bewertungen der einzelnen Störungen auszuwerten. Für die Bewertung der Störintensität wird das markierte Maß auf einen Wert α_{roh} im Bereich $0 \dots 1$ abgebildet. Die Bewertung der zeitlichen Eigenschaften mündet in der Feststellung, ob die Störung einen kurzzeitigen oder einen permanenten oder beide Charaktere aufweist. Zur Feststellung, mit welchen Anteilen die einzelnen Störungen die Qualitätsbewertung beeinflussen, erfolgt eine Normierung:

$$\alpha_i = \frac{\alpha_{roh,i}}{\sum_{k=1}^N \alpha_k} \quad (5.14)$$

In den weiteren Untersuchungsschritten wird das arithmetische Mittel der Anteile α_i aller Betrachter verwendet. Bei einem idealen 3D-Videoobjektqualitätsmaß sind die Bewertungsanteile α_i der subjektiven Bewertung und die Anteile $\tilde{\alpha}_i$ der objektiven Bewertung gleich groß. Daher müssen für die sukzessive Approximation die Koeffizienten β_i mittels der ermittelten Verteilung der Anteile α_i modifiziert werden. Hierfür wird von einer schrittweisen Mittelung ausgegangen, welche zu einem modifizierten Koeffizienten β^* bzw. Anteil $\tilde{\alpha}_i^*$ führt.

$$\tilde{\alpha}_i^* = \frac{R \cdot \tilde{\alpha}_i(\beta_i) + \alpha_i}{R + 1} \quad (5.15)$$

Der Faktor R beschreibt die Reife der bereits vorliegenden Koeffizienten β_i . Diese könnte durch die Anzahl der bisherigen Approximationsschritte beschrieben werden. Da nicht grundsätzlich davon auszugehen ist, dass diese Anzahl bekannt ist, muss die Reife als Bezugsgröße zu der aktuellen Testreihe angegeben werden. Hierfür sind folgende Einflussgrößen zu beachten:

- *Anzahl der Testobjekte*: Die Anzahl N der verfügbaren Objekte der aktuellen Testreihe.
- *Relation*: Der Faktor g_{rel} gibt das Verhältnis des Umfanges der aktuellen Testreihe zur Reife des vorliegenden Qualitätsmaßes an.
- *Repräsentativität*: Für das Messobjekt wird ein Wichtungsfaktor g_{repr} festgelegt, der für als typisch eingeschätzte Objekte gleich 1 und für weniger repräsentative Objekte entsprechend kleiner (im Bereich $0 \dots 1$) gesetzt wird.

Die Reife R ergibt sich damit zu:

$$R = \frac{N}{g_{rel} \cdot g_{repr}} \quad (5.16)$$

Im letzten Schritt können mittels des Ansatzes zur Mittelung (Gleichung 5.15) die modifizierten Koeffizienten β_i^* bestimmt werden:

$$\tilde{\alpha}_i^* = \frac{\beta_i^* \cdot p_{\psi,i}}{VQM} = \frac{\frac{N}{g_{rel} \cdot g_{repr}} \cdot \frac{\beta_i \cdot p_{\psi,i}}{VQM} + \alpha_i}{\frac{N}{g_{rel} \cdot g_{repr}} + 1} \quad (5.17)$$

$$\beta_i^* = \frac{\frac{N}{g_{rel} \cdot g_{repr}} \cdot \beta_i + \frac{VQM}{p_{\psi,i}} \cdot \alpha_i}{\frac{N}{g_{rel} \cdot g_{repr}} + 1} \quad (5.18)$$

Der Ansatz zur Mittelung ist für den Fall $p_{\psi,i} = 0$ nicht gültig. Dieser Fall tritt ein, wenn das Testobjekt bezüglich dieses Qualitätsaspektes ungestört ist, sodass dieses Testobjekt nicht zur Modifikation des Koeffizienten β_i^* dienen kann. Dieser Fall kann auch eintreten, wenn das mathematische Modell den Qualitätsaspekt nicht ausreichend beschreibt bzw. wenn der erweiterte subjektive Test einen abweichenden Aspekt erfasst. Da bei einer solchen Testreihe davon auszugehen ist, dass einige Werte für $p_{\psi,i}$ und β_i kontaminiert sind, führt ein winsorisiertes oder ein gestutztes Mittel zu einem sichereren Ergebnis. So können beispielsweise jeweils 5% der Werte für $p_{\psi,i}$ am oberen und unteren Ende unberücksichtigt bleiben.

Zusammenhang zwischen DMOS und 3DVQM

Die Bestimmung des Zusammenhanges zwischen dem DMOS und dem VQM erfolgt nach [ITU02], [VQEG03] durch eine nichtlineare Regression mit einer logistischen Funktion. Anschließend kann so der DMOS eines 3DVO mit dessen VQM prädiziert werden:

$$DMOS_{pred} = \frac{1}{1 + e^{(VQM - D_M) \cdot G}} \quad (5.19)$$

Die Bestimmung der Konstanten D_M und G kann mit Hilfe der Logit-Transformation erfolgen (siehe Unterabschnitt 2.3.7).

5.3 Exemplarische Bestimmung der Gewichtungskoeffizienten

Für den Nachweis der vorgeschlagenen Methodik zur Bestimmung der Gewichtungskoeffizienten wurde diese exemplarisch für die vorhandenen 3D-Videoobjekte und die bereits implementierten Qualitätsparameter durchgeführt.

5.3.1 Auswahl der Testobjekte

Für die Untersuchung wurden drei verschiedene 3D-Videoobjekte in je bis zu zehn Qualitätsstufen verwendet. Es handelt sich dabei um die Objekte „IAVAS“ (TU Ilmenau, [Due03], siehe Abbildung 4.6), „MPI-Dancer“ (MPI Saarbrücken, siehe Abbildung 4.7) und „Lara“ (ETH Zürich, Schweiz; siehe Abbildung 4.5). Alle drei 3D-Videoobjekte wurden mittels unterschiedlicher Verfahren erzeugt, sodass auch die auftretenden Störungen unterschiedlichen Charakter aufweisen:

1. „IAVAS“ Testobjekt: 1...10: Für die View-Synthesis der Testobjekte wurden Korrespondenzpunktpaare bzw. eine Kameraansicht ausgelassen.
2. „Lara“ Testobjekt: 11...13: Die Texturierung erfolgte mittels unterschiedlicher Datenraten.
3. „MPI-Dancer“ Testobjekt: 14...16: Es erfolgte eine unterschiedlich hohe Kompression der Texturdaten.

Für die 3D-Videoobjekte wurde ein typischer Navigationspfad (z. B. Umlauf auf Viertelkreis) oder ein typischer Betrachterstandpunkt ausgewählt. Um eine hohe Aufmerksamkeit

des Betrachters auf das Testobjekt zu ermöglichen, erfolgte die Präsentation nicht auf einem 3D-Terminal²⁷, sondern das 3D-Videoobjekt wurde dort aufgezeichnet und dem Betrachter als konventionelles Video vorgestellt. Die Sequenzen sind jeweils fünf bis zehn Sekunden lang. Das 3D-Videoobjekt „IAVAS“ wurde in einer Kulisse gezeigt, während die beiden anderen Objekte in einer einfarbigen Umgebung²⁸ gezeigt wurden.

5.3.2 Auswahl der zu bewertenden Störungen

Für die erweiterten subjektiven Tests mussten die Störungen ausgewählt werden, welche für die Testobjekte relevant sind und einen Bezug zu den Qualitätsparametern herstellen lassen.

- Lokale Störungen (Synthesefehler, Okklusionen etc.) bilden eine typische Störung der verwendeten 3D-Videoobjekte und werden mit Hilfe des Begriffes „fehlerhaft erscheinende Objektbereiche“ im Formular beschrieben.
- Konturstörungen bilden ebenfalls eine typische Störung von 3D-Videoobjekten und werden mit Hilfe des Begriffes „Konturfehler“ im Formular beschrieben.
- Ortsänderungen (Translationen) treten bei den Testobjekten geringfügig auf. Sie werden durch die Begriffe „Positionsfehler/Unruhe“ im Formular beschrieben.
- Kompressionsartefakte und andere, den Gesamteindruck schmälernde Störungen werden als „Unschärfe des gesamten Objektes“ beschrieben.
- Größenänderungen und Topologieänderungen sind bei den verwendeten Testobjekten kaum bzw. quasi nicht vorhanden und werden daher im Formular nicht genannt.

In Abbildung 5.5 ist das verwendete Formular für die erweiterten subjektiven Bewertungsreihen abgebildet.

5.3.3 Bestimmung der Initialwerte für die Koeffizienten

Entsprechend der in Unterabschnitt 5.2.2 vorgestellten Methodik müssen zunächst Initialwerte für die Gewichtungskoeffizienten β_i festgelegt werden. Hierfür wurden die Erfahrungen bei der Modellierung der Qualitätsparameter genutzt. In Tabelle 5.2 sind die Initialwerte der Koeffizienten für die in Unterabschnitt 5.1.4 gebildeten Qualitätsparameter sowie die dafür genutzten Grundlagen aufgelistet.

5.3.4 Bewertungsreihen

Für die erweiterten subjektiven Bewertungsreihen wurden sechs Betrachter mit einschlägigen Kenntnissen und normalen bzw. korrigierten Visus ausgewählt²⁹. Natürlich stellte die Bewertung eines neuen Medienobjektyps die Betrachter vor einige Schwierigkeiten:

- Die Referenzen haben bei weitem noch nicht die Qualität, welche allgemein als gut oder sehr gut empfunden wird. Sie haben ebenfalls Störungen, sodass die zusätzlichen Störungen des Testobjektes nur mit erhöhter Aufmerksamkeit zu erkennen sind.

²⁷z. B. ein VRML-Browser oder ein MPEG-4 3D-Player

²⁸nur Boden und Hintergrund

²⁹entsprechend der ITU-Empfehlung [ITU99]: 4...8 Betrachter

Qualitätsparameter $p_{\psi,i}$	Koeffizient β_i	Grundlage
$p_{position,m}$	5	Untersuchungen in Abschnitt 4.7
$p_{position,s}$	50	Untersuchungen in Abschnitt 4.7
$p_{position,a}$	1000	Untersuchungen in Abschnitt 4.7
$p_{groesse,m}$	1	Untersuchungen in Unterabschnitt 4.6.1
$p_{groesse,s}$	10	Untersuchungen in Unterabschnitt 4.6.1
$p_{groesse,a}$	100	Untersuchungen in Unterabschnitt 4.6.1
$p_{kontur,m}$	0,0005	Untersuchungen in Unterabschnitt 4.6.3
$p_{kontur,a}$	0,005	Untersuchungen in Unterabschnitt 4.6.3
$p_{topologie\#,m}$	1	Abschätzungen; Ergebnisse nach [Gel04]
$p_{topologie\#,s}$	20	Abschätzungen; Ergebnisse nach [Gel04]
$p_{topologie\Delta,m}$	0,2	Abschätzungen; Ergebnisse nach [Gel04]
$p_{topologie\Delta,s}$	2	Abschätzungen; Ergebnisse nach [Gel04]
$p_{lokal,m}$	5	Untersuchungen in Unterabschnitt 4.6.2
$p_{lokal,a}$	50	Untersuchungen in Unterabschnitt 4.6.2
$p_{global,m}$	5	Untersuchungen in Abschnitt 4.7
$p_{global,a}$	5	Untersuchungen in Abschnitt 4.7

Tabelle 5.2: Initialwerte für die Gewichtungskoeffizienten

- Den Betrachtern fiel die Bewertung der Qualität oft schwer, da sich auf Grund der fehlenden Erfahrungen noch kein ausgeprägtes Qualitätsempfinden entwickeln konnte.
- Die Termini zur Beschreibung der Störungen sind nicht selbsterklärend, da 3D-Videoobjekte und deren Störungsbeschreibungen noch nicht etabliert sind. Daher waren zusätzliche Erläuterungen für die Testpersonen erforderlich.

Die Ergebnisse der Testreihen wurden entsprechend dem in Unterabschnitt 5.2.2 vorgestellten Verfahren normiert. In Abbildung 5.6 sind die Bewertungen aller 16 Objekte dargestellt. Es ist sowohl der DMOS als auch das Intervall der aufgetretenen Bewertungen³⁰ dargestellt³¹. Wie zu erwarten, unterscheidet sich die Qualitätsbewertung der einzelnen Betrachter deutlich. Je mehr das zu bewertende Objekt vom Original abweicht, desto größer sind die Abweichungen in der Qualitätsbewertung. Dies ist auf die unterschiedlichen Erwartungshaltungen der Betrachter zurückzuführen. Wenn sich das zu bewertende Objekt kaum von der Referenz unterscheidet, gehen die Bewertungen jedoch kaum auseinander. Diese Tatsache spricht für die Verwendbarkeit der Testergebnisse.

In Abbildung 5.7 sind die Anteile α_i der einzelnen Bewertungsaspekte auf die Qualitätsbewertung dargestellt. Die Diagramme zeigen exemplarisch die Verteilung für vier Testobjekte. Ebenso wie im Diagramm für die Qualitätsbewertung werden hier jeweils der durchschnittliche Anteil sowie das Intervall aller Anteile von den sechs Bewertern dargestellt. Es wird deutlich, dass jeweils ein bis zwei Aspekte eines jeden Testobjektes die Qualitätsbewertung maßgeblich beeinflussten. Die sich durch die Testreihe ergebenden

³⁰Auf Grund der Randbedingungen (Bewertung für die Entwicklung eines Qualitätsmaßes; Test mit Experten; nur sechs Teilnehmer) wurde von einer statistischen Auswertung abgesehen.

³¹In der Auswertung wurde festgestellt, dass meistens eine Bewertung von den anderen deutlich abwich. Diese wurde für die Auswertung nicht genutzt.

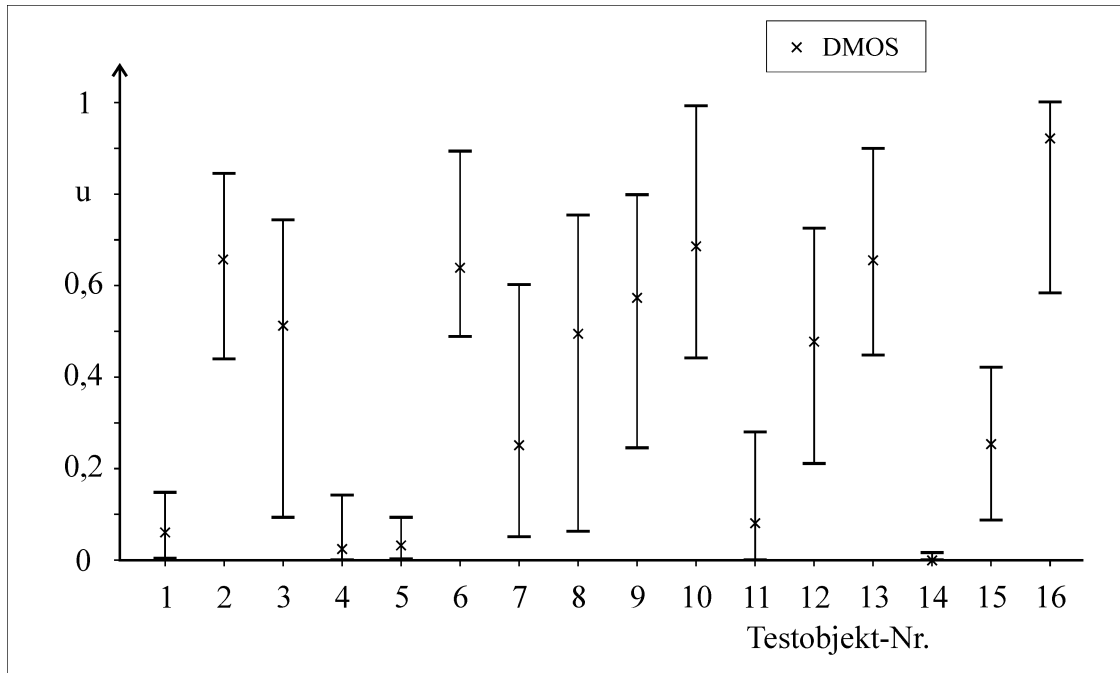


Abbildung 5.6: Subj. Qualitätsbewertung: DMOS und Intervall der Einzelbewertungen

Verteilungen der Aspekte sind bei den verschiedenen Testobjekten signifikant unterschiedlich. Diese Tatsache bestätigt, dass die Beschreibungen der Bewertungsaspekte bei den verschiedenen Betrachtern zu ähnlichen Assoziationen führte und daher ein Indiz für die Verwendbarkeit der Testergebnisse gegeben ist.

5.3.5 Objektive Bewertung

Für die Testobjekte wurden die in Unterabschnitt 5.1.4 vorgestellten 3D-Qualitätsparameter bestimmt und eine gewichtete Summierung (siehe Unterabschnitt 5.2.1) durchgeführt. In den Tabellen 5.3 und 5.4 sind für verschiedene Testobjekte die ermittelten Qualitätsparameter aufgelistet³². Im zweiten Schritt erfolgte die Wichtung der Qualitätsparameter auf Basis der für die Gewichtungskoeffizienten gewählten Initialwerte (siehe Tabelle 5.2).

5.3.6 Abbildung der objektiven Bewertung auf die subjektive Bewertung

Methodik

Für die Verifikation eines objektiven Qualitätsmaßes wird es auf den Wertebereich der subjektiven Bewertung abgebildet [ITU02], [VQEG03] u. a. Hierfür wird eine logistische Regression durchgeführt (siehe Unterabschnitte 2.3.7 und 5.2.2). Für die Rückführung auf eine lineare Regression wird folgende Substitution durch die Variable L vorgenommen:

$$DMOS_{pred} = \frac{1}{1 + e^{(VQM - D_M) \cdot G}}$$

$$L = \ln \left(\frac{1}{DMOS_{pred}} - 1 \right) = G \cdot VQM - G \cdot D_M \quad (5.20)$$

³²Die in Gleichung 4.34 verwendete Schwelle wurde für wa_{lokal} auf $c = 1, 2$ angepasst.

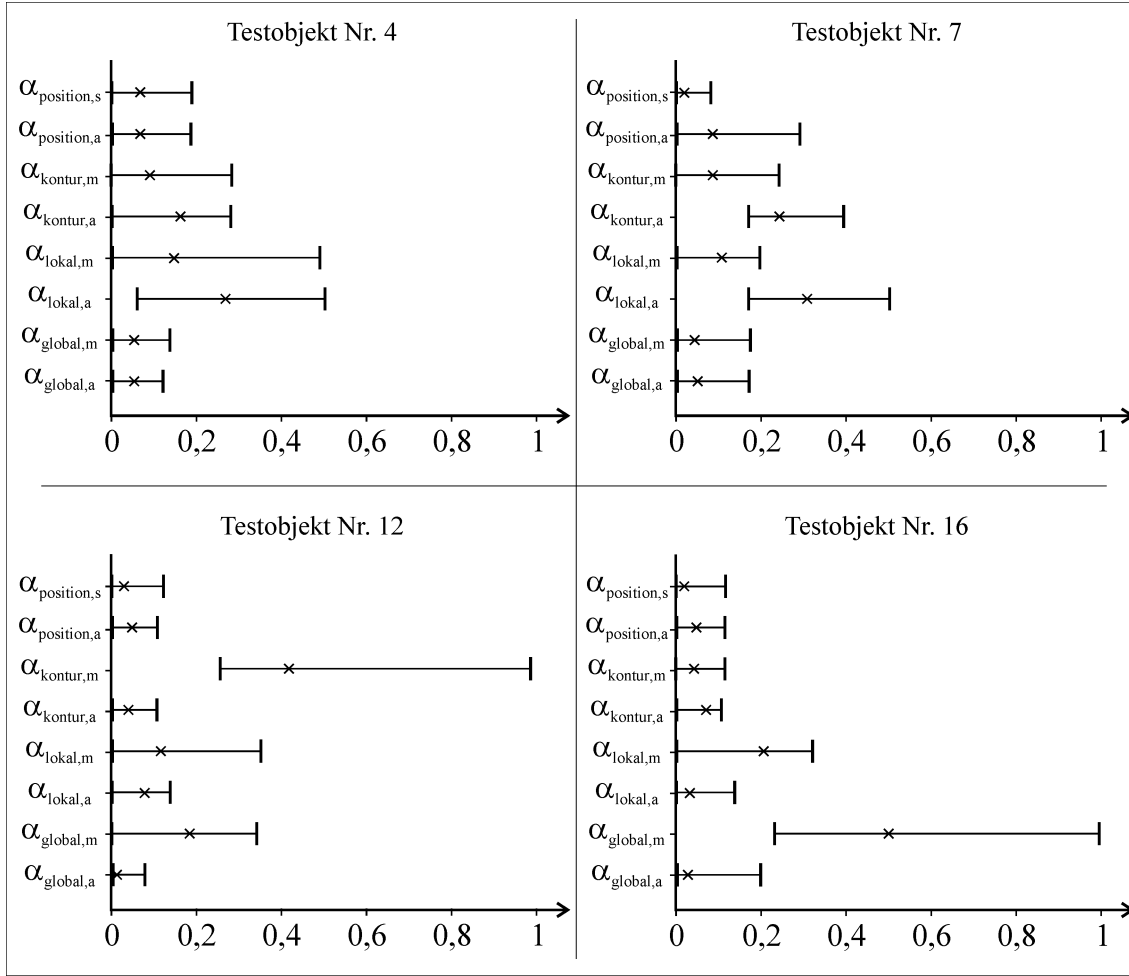


Abbildung 5.7: Anteile der subjektiven Bewertung für vier Testobjekte

Die Bestimmung der Regressionsvariablen erfolgt gemäß den Gleichungen 5.21 und 5.22.

$$G = \frac{\sum_{i=1}^N (VQM(i) - \overline{VQM}) \cdot (L(i) - \bar{L})}{\sum_{i=1}^N (VQM(i) - \overline{VQM})^2} \quad (5.21)$$

$$D_M = \overline{VQM} - \frac{1}{G} \cdot \bar{L} \quad (5.22)$$

Die Güte der Prädiktion wird nach [VQEG03] nach folgenden Kriterien beurteilt:

- *Genauigkeit der Prädiktion*: Größe der Abweichungen der Prädiktionen
- *Monotonieverhalten der Prädiktion*: Grad der Gleichrangigkeit zwischen den Prädiktionen und den subjektiven Bewertungsmaßen
- *Beständigkeit der Prädiktion*: Grad der Genauigkeit der Prädiktionen über eine Vielzahl an Testobjekten und Störungen hinweg

Als Maße für diese Kriterien werden verschiedene Korrelationskoeffizienten (Pearson, Spearman) sowie Fehlermaße (RMS u. ä.) genutzt.

	1	3	4	6	7	9	10
$p_{position,m}$	0	0	0	0	0	0	0
$p_{position,s}$	0,047	0,062	0,054	0,053	0,066	0	0
$p_{position,a}$	0,00093	0,0053	0,00063	0,00073	0,0019	0,0016	0,0017
$p_{groesse,m}$	0	0	0	0	0	0	0
$p_{groesse,s}$	0,089	0,11	0,051	0,083	0,11	0,16	0,15
$p_{groesse,a}$	0,0046	0,0097	0,0046	0,010	0,0090	0,016	0,013
$p_{kontur,m}$	8300	5900	6800	6400	7200	6700	7300
$p_{kontur,a}$	650	980	750	570	880	450	400
$p_{lokal,m}$	0,67	0,60	0,61	0,59	0,57	0,62	0,63
$p_{lokal,a}$	0,19	0,16	0,17	0,20	0,14	0,14	0,21
$p_{global,m}$	0,64	0,57	0,43	0,41	0,77	0,71	0,74
$p_{global,a}$	0,066	0,37	0,038	0,030	0,18	0,084	0,092
3DVQM	28,6	34,0	25,0	26,5	29,9	27,2	27,9

Tabelle 5.3: Qualitätsparameter für das 3DVO „IAVAS“

Neben einer Abbildung des objektiven Qualitätsmaßes auf die subjektive Qualitätsbewertung kann auf Grund der erweiterten subjektiven Testverfahren ein Vergleich der objektiven und subjektiven Bewertungsanteile vorgenommen werden. Damit kann eine Verifikation des 3D-Videoobjektqualitätsmaßes bereits auf der Ebene der Qualitätsparameter erfolgen.

Vergleich der objektiven und subjektiven Bewertungsanteile

In Abbildung 5.8 sind beispielhaft für vier Testobjekte die Anteile der subjektiven Bewertung α_i und die Anteile der objektiven Bewertung $\tilde{\alpha}_i$ gegenübergestellt. Die Diagramme zeigen eine gute Übereinstimmung zwischen objektiver und subjektiver Bewertung. Für die quantitative Erfassung der Übereinstimmung wird der Korrelationskoeffizient nach Pearson bestimmt³³. Dieser Korrelationskoeffizient ist jedoch nur für die Anteile α_i relevant, die einen signifikanten Einfluss auf die Qualitätsbewertung haben. In Tabelle 5.5 sind für die in dieser Testreihe verwendeten Qualitätsaspekte der Korrelationskoeffizient r sowie der mittlere Anteil $\bar{\alpha}$ der subjektiven Bewertung aufgelistet. Die Tabelle verdeutlicht, dass für die Anteile der signifikanten Qualitätsaspekte (z. B. $\alpha_{lokal,a}$, $\alpha_{global,m}$) eine hohe Korrelation von $> 0,8$ erzielt wurde. Lediglich für einen signifikanten Aspekt ($\alpha_{kontur,m}$) wurde eine negative Korrelation erreicht.

- *Conclusio 5.9:* Die Korrelation zwischen den Anteilen der objektiven und subjektiven Bewertung ist für die meisten Qualitätsaspekte bereits ausreichend hoch, um diese zu einem objektiven Qualitätsmaß zusammenführen zu können.

Vergleich zwischen der objektiven und der subjektiven Qualitätsbewertung

Für die Testreihe ergibt sich gemäß den Gleichungen 5.21 und 5.22 folgende Abbildung der objektiven Bewertung auf die subjektive Bewertung:

$$DMOS_{pred} = \frac{1}{1 + e^{(VQM - 41,65) \cdot (-0,0440)}} \quad (5.23)$$

³³Das Merkmal -Anteil α_i - wird als intervallskaliert eingestuft.

	11	12	13	14	15	16
$p_{position,m}$	0	0	0	0	0	0
$p_{position,s}$	0	0	0	0	0	0
$p_{position,a}$	0,00002	0,0004	0,0007	0,00003	0,00003	0,0003
$p_{groesse,m}$	0	0	0	0	0	0
$p_{groesse,s}$	0,10	0,17	0,13	0	0	0
$p_{groesse,a}$	0,021	0,034	0,014	0,00084	0,00084	0,00084
$p_{kontur,m}$	820	1200	1600	54	54	54
$p_{kontur,a}$	74	82	0	10	10	10
$p_{lokal,m}$	0,68	0,68	0,64	0,78	0,76	0,71
$p_{lokal,a}$	0,072	0,054	0,033	0,018	0,0074	0,0040
$p_{global,m}$	0,58	0,85	1,3	0,93	1,6	1,3
$p_{global,a}$	0,023	0,027	0	0	0	0
3DVQM	14,1	17,0	15,8	9,62	12,3	10,3

Tabelle 5.4: Qualitätsparameter für die 3DVO „Lara“ und „MPI-Dancer“

	r	$\bar{\alpha}$
$\alpha_{position,s}$	0,30	0,046
$\alpha_{position,a}$	0,33	0,091
$\alpha_{kontur,m}$	-0,22	0,12
$\alpha_{kontur,a}$	0,76	0,14
$\alpha_{lokal,m}$	0,29	0,15
$\alpha_{lokal,a}$	0,84	0,21
$\alpha_{global,m}$	0,88	0,16
$\alpha_{global,a}$	-0,05	0,057

Tabelle 5.5: Korrelation zwischen den objektiven und subjektiven Bewertungsanteilen

In Abbildung 5.9 sind für die 16 Objekte der Testreihe die prädizierten Werte für den DMOS und die ermittelten Werte für den DMOS in einem Streudiagramm (*scatter diagram*) dargestellt. Im Idealfall lägen sämtliche Wertepaare auf der Geraden $DMOS = DMOS_{pred}$. Da sich das 3D-Videoobjektqualitätsmaß in einem frühen Entwicklungsstadium befindet, ist die Güte seiner Prädiktion noch nicht hoch³⁴. Aus diesem Grund ist es sinnvoll, zunächst nur die Gleichrangigkeit der Prädiktionen als Kriterium zu nutzen. Diese kann mit dem Spearman-Rangkorrelationskoeffizienten r_s gemessen werden:

$$r_s = 1 - 6 \cdot \sum_{i=1}^N \frac{d^2}{N \cdot (N-1)^2} \quad (5.24)$$

Für die Testreihe ergibt sich der Rangkorrelationskoeffizient zu $r_s = 0,08$. Diese Korrelation ist noch nicht groß genug, um eine Anwendung des Qualitätsmaßes zu ermöglichen. Die Ursachenanalyse führte neben einer unangepassten Gewichtung und unausgereiften Parametermodellen zu folgenden Problemen:

- Beispielsweise weist das Testobjekt Nr. 16 sehr starke Artefakte auf Grund der hohen Kompression der Texturdaten auf und wird daher subjektiv als sehr schlecht

³⁴Dies wird durch die große Streuung der Werte im Streudiagramm deutlich.

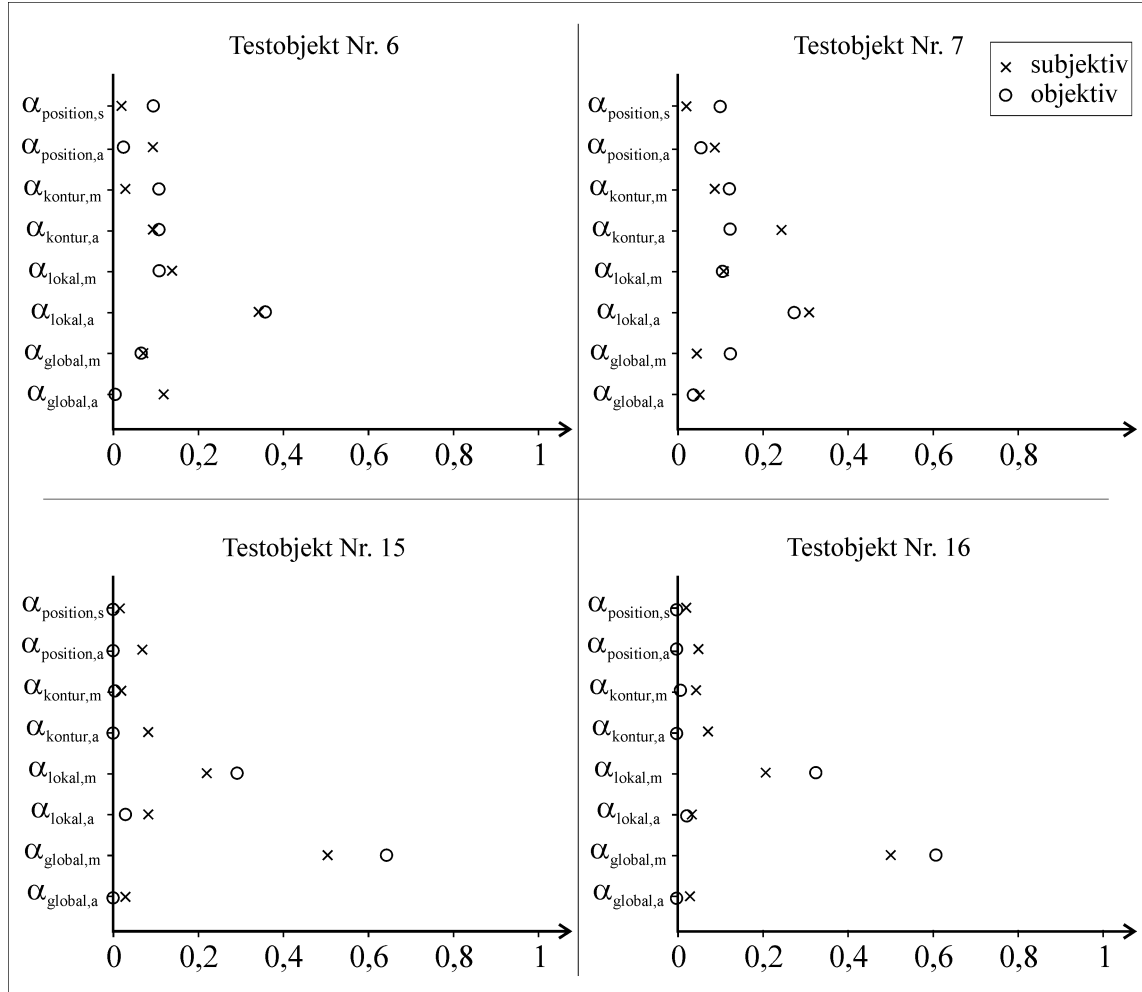


Abbildung 5.8: Anteile α_i und $\tilde{\alpha}_i$ der subjektiven und der objektiven Bewertung

bewertet. Diese Artefakte werden mit dem Parameter $p_{\text{global,m}} = 1,3$ erfasst. Da das Objekt aber kaum andere Störungen aufweist, ist das 3DVQM vergleichsweise gut, sodass die objektive Bewertung nicht mit der subjektiven Bewertung korreliert.

- Einige Qualitätsparameter repräsentieren mehrere Qualitätsaspekte oder Störgrößen. Der Einfluss der verschiedenen Störungen auf die Wahrnehmung ist jedoch gerade im überschwelligem Bereich³⁵ unterschiedlich. Da in dieser Testreihe verschiedenartige, deutlich apparente Störungen auftraten, ist die Auswirkung auf die Qualitätsparameter abweichend von der subjektiven Bewertung. Beispielsweise ist der hohe Wert für $p_{\text{lokal,a}}$ bei Testobjekt Nr. 1 durch Verzerrungen und bei Testobjekt Nr. 6 durch Okklusionen bedingt, während Objekt Nr. 1 subjektiv deutlich besser bewertet wird.

Schränkt man jedoch die Vielfältigkeit der Störungen ein, steigt erwartungsgemäß die Korrelation an. Diese beträgt für die Testobjekte Nr. 1...10 bereits $r_s = 0,23$ für die Initialwerte der Gewichtungskoeffizienten.

- *Conclusio 5.10:* Die Korrelation zwischen der objektiven und der subjektiven Qualitätsbewertung ist bei einer Verwendung von Objekten mit verschiedenartigen und stark überschwelligem Störungen sehr klein.

³⁵d. h. weit über der Wahrnehmungsschwelle

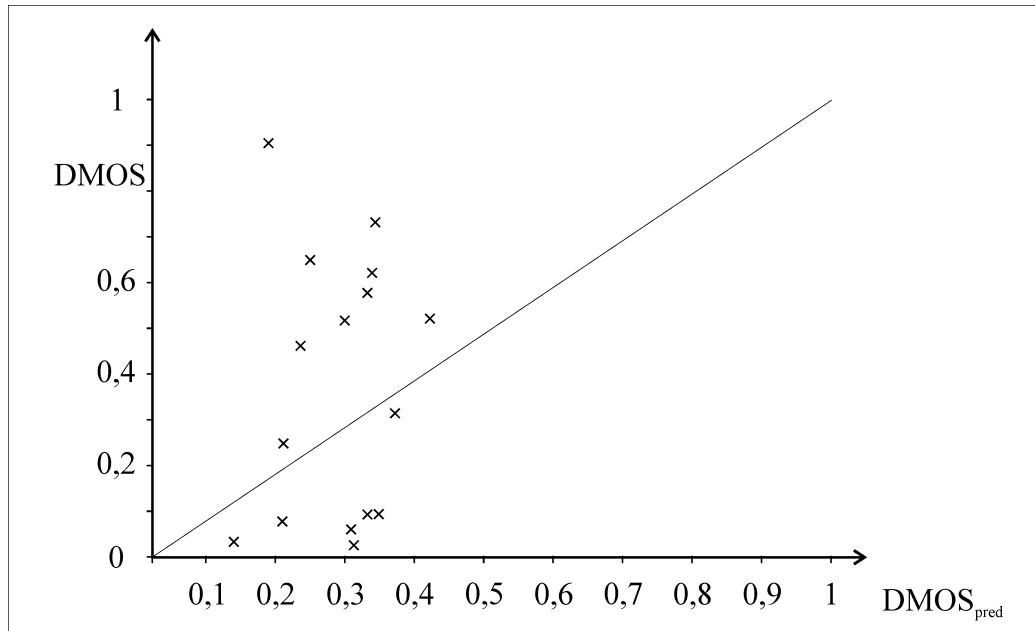


Abbildung 5.9: Streudiagramm für den DMOS und dessen Prädiktion

Dieses Problem spiegelt die Besonderheiten der subjektiven Bewertung der untersuchten Objekte wider. Wenn zwei Objekte durch jeweils eine Störung stark beeinträchtigt werden, so ist die subjektive Bewertung der resultierenden Qualität individuell sehr verschieden³⁶. Damit kann die subjektive Bewertung mit nur einer geringen Korrelation mit Hilfe eines objektiven Einzelmaßes prädiziert werden.

5.3.7 Modifikation der Gewichtungskoeffizienten

Im nächsten Schritt erfolgt die Modifikation der Gewichtungskoeffizienten β_i entsprechend der in Unterabschnitt 5.2.2 vorgeschlagenen Methode. Dies erfordert zunächst die Bestimmung der Einflussgrößen. Die Testreihe umfasst $N = 16$ Testobjekte. Die Relation g_{rel} ist in Bezug auf die Initialwerte nicht im originären Sinne bestimmbar. Die Initialwerte wurden anhand von Erfahrungswerten bestimmt. Insofern wird $g_{rel} = 1$ gesetzt. Für die Bestimmung der Repräsentativität g_{repr} ist zu beachten, dass die verwendeten 3D-Videoobjekte keine hohe Repräsentativität besitzen, da sich noch kein 3DVO-Verfahren etabliert hat. Daher wird $g_{repr} = 0,5$ gesetzt.

$$R = \frac{N}{g_{rel} \cdot g_{repr}} = \frac{16}{1 \cdot 0,5} = 32 \quad (5.25)$$

Da die Ergebnisse dieser Testreihe für die erste Approximation dienen, werden unmittelbar die Initialwerte (siehe Tabelle 5.2) modifiziert. In Tabelle 5.6 sind die modifizierten Gewichtungskoeffizienten den Initialwerten gegenübergestellt³⁷.

³⁶Beispielsweise wurde das Testobjekt Nr. 6 (deutlich apparente Okklusionen) von einem Bewerter mit einer Differenz von 40 bewertet, während er das Testobjekt Nr. 16 (hohe Texturkompression) mit 25 bewertete. Ein anderer Bewerter gab für Objekt Nr. 6 eine Differenz von 2 und für Nr. 16 eine Differenz von 40 an.

³⁷Für die Koeffizienten β_i der Parameter $p_{groesse,m}$, $p_{groesse,s}$, $p_{groesse,a}$ und $p_{position,m}$ erfolgte keine Modifikation, da die zugehörigen Merkmale in der Testreihe kaum oder gar nicht vorhanden waren.

Qualitätsparameter $p_{\psi,i}$	Koeffizient β_i	β_i^*
$p_{position,m}$	5	5
$p_{position,s}$	50	21,8
$p_{position,a}$	500	1494
$p_{groesse,m}$	1	1
$p_{groesse,s}$	10	10
$p_{groesse,a}$	100	100
$p_{kontur,m}$	0,0005	0,0032
$p_{kontur,a}$	0,005	0,021
$p_{lokal,m}$	5	5,03
$p_{lokal,a}$	50	69
$p_{global,m}$	5	4,44
$p_{global,a}$	5	36,18

Tabelle 5.6: Initialwerte und modifizierte Werte der Gewichtungskoeffizienten

5.3.8 Auswertung der 3DVQM-Approximation

Mit Hilfe der in Unterabschnitt 5.3.7 ermittelten Gewichtungskoeffizienten wurden die Werte des modifizierten 3DVQM für die 16 Testobjekte bestimmt. Zunächst erfolgt ein Vergleich der Anteile der subjektiven und der objektiven Bewertung. In Tabelle 5.7 sind die Werte der Korrelation auf Basis der modifizierten Gewichtungskoeffizienten mit den Initialwerten vergleichend gegenübergestellt. Die Korrelation der einzelnen Anteile änderte sich geringfügig.

	r für β_i	r für β_i^*
$\alpha_{position,s}$	0,30	0,29
$\alpha_{position,a}$	0,33	0,28
$\alpha_{kontur,m}$	-0,22	-0,08
$\alpha_{kontur,a}$	0,76	0,76
$\alpha_{lokal,m}$	0,29	0,33
$\alpha_{lokal,a}$	0,84	0,56
$\alpha_{global,m}$	0,88	0,88
$\alpha_{global,a}$	-0,05	0,10

Tabelle 5.7: Korrelation zwischen den objektiven und subjektiven Bewertungsanteilen nach der Modifikation der Gewichtungskoeffizienten

Im zweiten Schritt erfolgte wiederum eine Abbildung der objektiven auf die subjektive Qualitätsbewertung. Die optimierte Prädiktion für den DMOS ergibt sich nach Gleichung 5.26.

$$DMOS_{pred} = \frac{1}{1 + e^{(VQM - 53,21) \cdot (-0,0423)}} \quad (5.26)$$

Letztlich wurde die Korrelation zwischen der objektiven und der subjektiven Qualitätsbewertung bestimmt. Im ersten Fall wurde diese wieder für die gesamte Testreihe bestimmt. In Abbildung 5.10 ist das Streudiagramm für alle Prädiktionen dargestellt. Erwartungsgemäß blieb diese auf Grund der beschriebenen Ursachen mit $r_s = 0,07$ sehr klein. Im

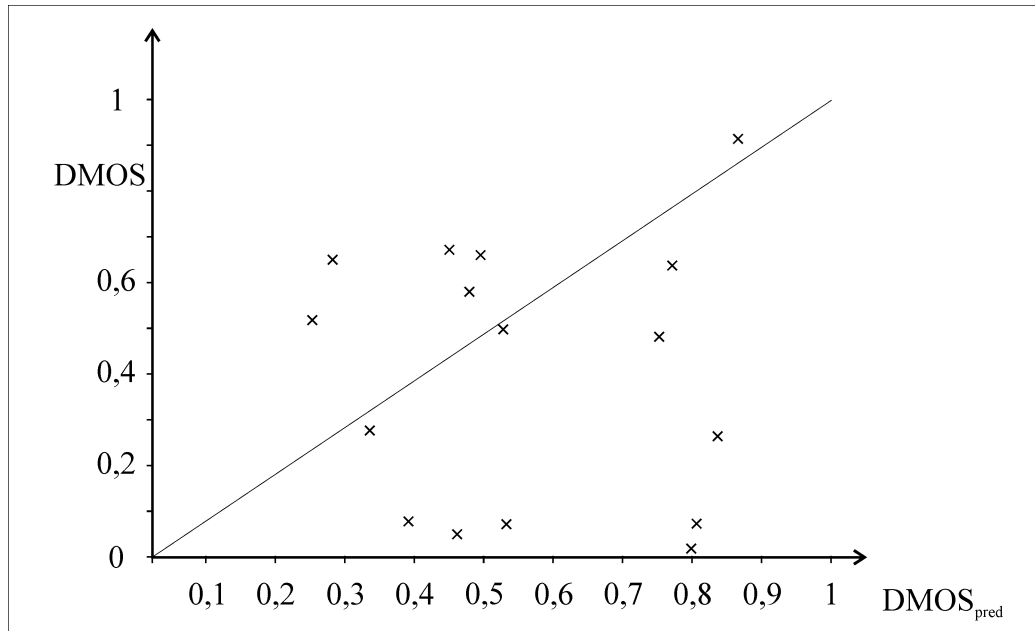


Abbildung 5.10: Streudiagramm für den DMOS und dessen optimierte Prädiktion

zweiten Fall wurde die Korrelation für die Testobjekte Nr. 1...10 bestimmt. Diese verbesserte sich durch die optimierten Gewichtungskoeffizienten von $r_s = 0,23$ auf $r_s = 0,45$ (siehe Tabelle 5.8).

	r_s für β_i	r_s für β_i^*
Testobjekte Nr. 1...10	0,23	0,45

Tabelle 5.8: Verbesserung der Korrelation nach der Optimierung der Gewichtungskoeffizienten

5.4 Validierung der Methodik

Im Abschnitt 5.3 wurde die vorgeschlagene Methodik zur Bildung eines 3D-Videoobjektqualitätsmaßes anhand der zur Verfügung stehenden Testobjekte exemplarisch angewendet. Die Durchführung der subjektiven und objektiven Bewertungsverfahren erlaubt die Validierung der Methodik und zeigt deren Grenzen auf.

Ein wesentlicher Bestandteil der Methodik sind die erweiterten subjektiven Testverfahren. Ihre Durchführung (siehe Unterabschnitt 5.3.4) und ihre Auswertung zeigen, dass sie für die Ermittlung einer subjektiven Qualitätsbewertung und der hierfür signifikanten Aspekte geeignet ist. Trotz fehlender weitreichender Erfahrungen in der Nutzung solcher Objekte führte die Auswertung zu verwertbaren Erkenntnissen (siehe u. a. Abbildung 5.7). Nach den Aussagen der Testpersonen erforderten die Testreihen eine hohe Konzentration und führten zu einer hohen Anspannung. Es ist wichtig, den Testpersonen insbesondere die Termini zur Benennung der Qualitätsaspekte und Störungen zu erläutern. Ebenso ist es notwendig, die Termini anzupassen, wenn eine Weiterentwicklung der mathematischen Modelle für die Qualitätsparameter erfolgt ist. Ein Nachteil des vorgeschlagenen Verfahrens ist, dass es den interaktiven Charakter der Nutzung von 3D-Videoobjekten nicht

berücksichtigen kann³⁸.

- *Conclusio 5.11*: Das vorgeschlagene Verfahren zur erweiterten subjektiven Qualitätsbewertung ist geeignet, die für die Methodik notwendigen Bewertungen zu ermitteln. Das Verfahren bedarf grundsätzlich einer Adaption an künftige Testreihen.

Ein erster Schritt zur Validierung der Methodik und des vorgeschlagenen Qualitätsmaßes ist der Vergleich der objektiven mit den subjektiven Bewertungsanteilen. Der Pearson-Korrelationskoeffizient ist für die wichtigsten Bewertungsanteile größer als 0,8. Damit kann auf eine prinzipielle Eignung der mathematischen Modelle der Qualitätsparameter und deren Gewichtungskoeffizienten geschlossen werden. Die Korrelation zwischen der objektiven und subjektiven Qualitätsbewertung in Form eines einzelnen Maßes verschwindet, wenn die Art der Störungen sehr unterschiedlich ist und diese stark überschwellig sind. Wird eine sinnvolle Beschränkung der Störungen bzw. der Videoobjekte vorgenommen, so wird auch für das 3D-Videoobjektqualitätsmaß eine Anwendbarkeit erzielt. Der Korrelationskoeffizient betrug für die Initialwerte der Gewichtungskoeffizienten bereits 0,23 innerhalb einer Klasse von 3D-Videoobjekten.

- *Conclusio 5.12*: Die mathematischen Modelle der wichtigsten Qualitätsparameter und deren Gewichtungskoeffizienten sind für die Bildung eines 3D-Videoobjektqualitätsmaßes geeignet.

Der zweite wesentliche Bestandteil der Methodik ist die Modifikation der Gewichtungskoeffizienten. Diese Modifikation führte nicht zur Verbesserung der Prädiktion der Qualitätsbewertung der gesamten Testreihe, da die Problematik der Unterschiedlichkeit der Störgrößen dies verhindert. Jedoch führte die Modifikation zu einer Verbesserung der Prädiktion innerhalb einer Klasse von 3D-Videoobjekten. So konnte beispielsweise der Korrelationskoeffizient von 0,23 auf 0,45 verbessert werden.

- *Conclusio 5.13*: Die Methode der Approximation der Gewichtungskoeffizienten β_i auf Basis erweiterter subjektiver Testreihen führt bei der Verwendung geeigneter Ausgangswerte zu einer Verbesserung der Prädiktion des DMOS.

Der Verbesserung der Korrelation sind Grenzen gesetzt, welche jedoch nicht durch die Methodik begründet sind. Der wichtigste Grund hierfür sind die Grenzen der mathematischen Modelle der Qualitätsparameter zur Erfassung von Qualitätsaspekten und Störgrößen.

- *Conclusio 5.14*: Eine Approximation der Gewichtungskoeffizienten β_i ist nur insoweit sinnvoll, wie die Qualitätsparameter eine Erfassung der Qualitätsmerkmale und Störgrößen erlauben.

³⁸Dies liegt darin begründet, dass die Präsentation einer Referenz und eines Testobjektes erfolgt (siehe Unterabschnitt 5.2.2). Für eine interaktive subjektive Bewertung ist eine Single-Stimulus-Methode zu verwenden.

6 Ergebnisse und Ausblick

6.1 Zusammenfassung der Ergebnisse

In der vorliegenden Arbeit wurden Methoden und Verfahren zur Qualitätsbewertung von natürlichen 3D-Videoobjekten vorgestellt und exemplarisch verifiziert. Bevor jedoch ein Modell der Qualität solcher Objekte geschaffen werden konnte, waren verschiedene grundsätzliche Betrachtungen zu diesem neuen Medienobjekttyp notwendig. Ausgangspunkt war ein Überblick über die unterschiedlichen Entwicklungen, welche das gemeinsame Ziel haben, visuelle Medienobjekte um die dritte Dimension zu ergänzen. Damit konnte eine Definition des Begriffes *3D-Videoobjekt* erfolgen. Ein besonderes Merkmal dieses Medienobjekttyps ist es, dass er auf unterschiedliche Weise generiert werden kann. Für eine nähere Untersuchung von 3DVO-Merkmalen war selbstverständlich eine Auseinandersetzung mit den mannigfaltigen Generierungsverfahren nötig, welche ein Spektrum von bildbasierten bis hin zu modellbasierten Verfahren bilden. Da sich dieser Medienobjekttyp noch in einem frühen Entwicklungsstadium befindet, sind weitere und verbesserte Verfahren zu erwarten. Außerdem ist davon auszugehen, dass sich auf Grund der unterschiedlichen Anwendungen für 3D-Videoobjekte auch mehrere Verfahren für deren Generierung etablieren werden.

Eine wichtige Anforderung an die zu entwickelnde Methodik zur Qualitätsbewertung war die Unabhängigkeit von dem jeweiligen Verfahren, welches zur Generierung des 3D-Videoobjektes verwendet wird. Diese Anforderung mündete in einer Methode, bei der die 3D-Videoobjekte auf der Ebene der generierten Ansichten bewertet werden. Damit wurde gleichzeitig erreicht, dass die gesamte Kette von der Aufnahme bis zur Darstellung der Videoobjekte zum Gegenstand der Bewertung geworden ist¹. In diesem Punkt hebt sich diese Methode von anderen Ansätzen ab, welche beispielsweise der Fragestellung nachgehen: Wie lässt sich anhand eines texturierten 3D-Modells die Qualität der sich damit ergebenden Ansichten ermitteln?

Für die Qualitätsbewertung musste vorab die zentrale Frage nach der Referenz beantwortet werden. Die Untersuchungen führten zu dem Schluss, dass ein 3D-Videoobjekt mit einem *idealen* 3D-Videoobjekt verglichen werden muss. Jedoch ist das Attribut *ideal* auf die gegebenen Voraussetzungen zu beziehen. Mit diesem Attribut ist hier nicht gemeint, dass die Mehrzahl der Betrachter das Objekt als *ideal* bewerten, sondern es ist vielmehr damit gemeint, dass das Objekt im Rahmen der technischen Möglichkeiten und in Bezug auf die geplante Anwendung das Optimum darstellt. Aber auch für den Terminus *Referenz* sind spezielle Eigenschaften zu berücksichtigen. So kann beispielsweise bereits bei einem kleinen Blickwinkelfehler das 3D-Videoobjekt einige Bereiche des Aufnahmeobjektes zeigen, welche die Referenz nicht zeigt.

3D-Videoobjekte sind im Unterschied zu konventionellem Video zu jedem Zeitpunkt von der Blickpunktwahl abhängig. Diese Eigenschaft führte zu der Frage, welche Repräsentation für die Qualitätsbewertung verwendet werden kann. Das Ergebnis der Untersuchungen ist, dass jeweils die Ansichten einer typischen Navigation bzw. Blickpunktwahl zu ver-

¹Damit können die Auswirkungen von Veränderungen innerhalb der gesamten Kette auf die Qualität ermittelt werden.

wenden sind. Dies ist beispielsweise vergleichbar mit der Methode zur Bewertung eines Videoencoders. Diesen kam man unmöglich mit jedem potenziell auftretenden Videomaterial überprüfen. Daher verwendet man in gleicher Weise repräsentatives Testmaterial.

Eine ausgiebige Untersuchung wurde den Qualitätsmerkmalen von 3D-Videoobjekten gewidmet. Diese Objekte weisen zum Teil völlig neuartige Eigenschaften im Vergleich zu konventionellem Video auf. Es können Blickwinkelfehler, Größenänderungen, Okklusionen, Silhouettenverzerrungen etc. auftreten. Diese Störungen führen dazu, dass für den Vergleich kein Bildpunktbezug vorausgesetzt werden kann. Daher wird in der vorgeschlagenen Methode zur Bestimmung der Qualitätsparameter zunächst eine Erkennung von Skalierungs- und Translationsfehlern vorgenommen. Hierfür werden das hierarchische Block-Matching und die eindimensionale DFT der Konturfunktion angewendet. Die Ergebnisse dienen als Qualitätsparameter und zur Korrektur dieser Fehler. Nach dieser Korrektur ist ein statistischer Bildpunktbezug möglich. Hier kommen die Bestimmung des örtlichen Informationsgehaltes und der Histogrammvergleich zur Anwendung. Einige Eigenschaften lassen sich mittels globaler statistischer Größen und andere mit Hilfe eines lokalen Bezuges dieser Größen erfassen. Beispielsweise lassen sich Okklusionen und verfahrensspezifische Verzerrungen auf Basis der lokalen Größen ermitteln.

3D-Videoobjekte weisen im Gegensatz zu konventionellem Video eine Silhouette auf. Die Silhouette wird in der errechneten Ansicht als Transparenzinformation mit Hilfe eines vierten Kanals repräsentiert. Die Silhouette kann spezielle Störungen aufweisen, welche als Qualitätsaspekte in Erscheinung treten. Silhouettenverzerrungen können mit Hilfe der eindimensionalen DFT der Konturfunktion erkannt werden. Andererseits dient die Silhouette zur Erkennung weiterer Störungen (z. B. Größenfehler). Die Silhouette kann auch durch Topologieänderungen gestört sein. Beispielsweise können Öffnungen in der Silhouette entstehen oder korrekte Öffnungen verschwinden.

Mit Hilfe der untersuchten Methoden wurden mathematische Modelle für die Qualitätsmerkmale entwickelt. Für einige Merkmale (z. B. Translationsfehler) erwies sich der Einsatz mehrere Methoden für notwendig. Dahingegen müssen andere Merkmalsklassen indirekt über ihre statistischen Auswirkungen erkannt werden. Beispielsweise lassen sich Okklusionen und Blickwinkelfehler wegen der Referenzfrage nur undifferenziert feststellen.

Für die Qualitätsmerkmale wurden im nächsten Schritt die dynamischen Eigenschaften modelliert. Die dynamischen Eigenschaften lassen sich bei 3D-Videoobjekten in zeitabhängige und blickpunktsänderungsabhängige Eigenschaften unterscheiden. Da eine Unterscheidung dieser Eigenschaften nur teilweise anhand der berechneten Ansichten möglich ist, werden Blickpunktsänderungen mit Hilfe des Navigationspfades² bestimmt. Diese müssen ohnehin gegeben sein, da sie auch für die Verwendung des 3D-Videoobjektes in einer 3D-Anwendung erforderlich sind. Blickpunkts- und Blickrichtungsänderungen infolge von Navigationen beeinflussen die Wahrnehmung des 3D-Videoobjektes und beeinflussen daher die Qualitätsbewertung.

Auf Basis der Merkmalsmodelle wurden Qualitätsparameter entwickelt. Diese berücksichtigen bereits verschiedene Aspekte der visuellen Wahrnehmung. Um diese Aspekte mathematisch zu beschreiben, werden Skalierungen, Begrenzungsfunktionen, Abstandsmaße und die Koeffizienten der Blickpunktsänderung verwendet. Die in dieser Arbeit entwickelten Qualitätsparameter stellen eine Auswahl dar und können beliebig ergänzt und modifiziert werden.

Die vorgeschlagene Methode zur Bildung eines 3D-Videoobjektqualitätsmaßes sieht in

²ggf. mit Hilfe der Kameradaten einer virtuellen Kamera

ihrem letzten Schritt die gewichtete Zusammenführung zu einem Einzelmaß der Qualität vor. Diese Verknüpfung ist charakteristisch für Faktorenmodelle und erfolgt beispielsweise mit Hilfe der multiplen Regression [Fec00]. Da die dafür notwendigen Voraussetzungen im Falle der hier untersuchten 3D-Videoobjekte und deren Qualitätsparameter nicht vollständig gegeben sind (siehe Unterabschnitt 5.2.2), wurde in dieser Arbeit eine neue Methodik vorgeschlagen und exemplarisch angewendet. Diese Methodik sieht eine schrittweise Anpassung der Gewichtungskoeffizienten auf Basis erweiterter subjektiver Bewertungsreihen vor. Damit werden folgende Anforderungen erfüllt:

- Das 3D-Videoobjektqualitätsmaß ist bereits für erste Untersuchungen zur 3D-Videoobjektqualitätsbewertung verfügbar und kann schrittweise verbessert werden.
- Die Methodik erlaubt eine Anpassung an künftige 3DVO-Verfahren, deren spezifische Qualitätsaspekte und Störungen.
- Optimierte und neue Qualitätsparameter lassen sich problemlos innerhalb dieser Methodik verwenden.
- Das 3D-Videoobjektqualitätsmaß lässt sich an eine ausgewählte Klasse von 3D-Videoobjekten (z. B. bildbasiert) anpassen. Dadurch können bessere Prädiktionen der subjektiven Bewertung erzielt werden.

In den erweiterten subjektiven Bewertungsreihen entscheiden die Testpersonen sowohl über den gesamten Qualitätseindruck als auch über den Einfluss der einzelnen Aspekte. Für den Ablauf der Bewertungsreihe und die zu verwendende Qualitätsskala wurde auf Erfahrungen aus den ITU-Standards zur subjektiven Bewertung von Bewegtbild zurückgegriffen. Unter Berücksichtigung der besonderen Voraussetzungen wird für den Ablauf die DSIS-Methode und die Bewertungsskala der DSCQS-Methode verwendet. Die Bewertung der Qualitätsaspekte und deren dynamischen Eigenschaften erfolgt auf separaten Skalen mit einer größeren Einteilung.

Die Optimierung der Gewichtungskoeffizienten auf Basis der Ergebnisse der erweiterten subjektiven Bewertungsreihen orientiert sich an der Reife des aktuellen 3D-Videoobjektqualitätsmaßes. Zu diesem Zweck wurde die Repräsentativität der Testreihe und deren Relation zum aktuellen Qualitätsmaß bestimmt.

Die vorgeschlagene Methodik zur Bildung eines 3D-Videoobjektqualitätsmaßes wurde für die zur Verfügung stehenden 3D-Videoobjekte exemplarisch durchgeführt. Die erweiterten subjektiven Testreihen führten zu auswertbaren Ergebnissen. Der Einfluss der einzelnen Qualitätsaspekte auf die Bewertung konnte klar ermittelt werden. Nach den Aussagen der Testpersonen ist es wichtig, dass die Besonderheiten der 3D-Videoobjektbewertung³ und die Termini zur Beschreibung der Qualitätsaspekte eingehend erläutert werden.

Die Korrelation zwischen der objektiven und der subjektiven Bewertung der Qualitätsaspekte war bereits für die erste Testreihe ausreichend hoch. Sie betrug für die wichtigsten Aspekte mehr als 0,8. Damit ist eine prinzipielle Eignung der Qualitätsparameter und der Methodik nachgewiesen. Für den Vergleich der objektiven mit der subjektiven Qualitätsbewertung wurden diese mit Hilfe der Logit-Transformation aufeinander abgebildet. Die Korrelation zwischen der objektiven und der subjektiven Qualitätsbewertung über die gesamte Testreihe war sehr klein. Dies hat folgende Gründe:

³insbesondere die Idealität der Referenz

- Die Objekte der Testreihe wiesen unterschiedliche Störungen auf, die zum Teil stark überschwellig waren (z. B. Verzerrungen, Okklusionen, Texturdatenkompressionsartefakte). Diese werden bei der subjektiven Bewertung im Vergleich sehr individuell beurteilt. Daher kann ein objektives Einzelmaß grundsätzlich keine sehr gute Prädiktion ermöglichen.
- Einige Qualitätsparameter repräsentieren mehrere Qualitätsaspekte und Störungen. Insbesondere bei überschwelligen Störungen ist deren Linearität untereinander nicht mehr gegeben.
- Die Qualitätsparameter und deren Gewichtungskoeffizienten sind noch ungenau und bedürfen Verbesserungen.

Bei einer Begrenzung der Vielfalt der 3D-Videoobjekttypen stieg die Korrelation zwischen der objektiven und der subjektiven Qualitätsbewertung deutlich an.

Im letzten Schritt erfolgte die Modifikation der Gewichtungskoeffizienten auf Basis der Ergebnisse der ersten Testreihe. Durch die Anwendung des optimierten 3DVQM konnte der Spearman-Rangkorrelationskoeffizient für eine 3DVO-Klasse von 0,23 auf 0,45 verbessert werden. Damit konnte die Anwendbarkeit der vorgeschlagenen Methodik exemplarisch nachgewiesen werden.

6.2 Künftige Arbeiten

Aus der vorliegenden Arbeit ergeben sich viele Ansatzpunkte und Fragestellungen für künftige Arbeiten. Dies liegt selbstverständlich auch darin begründet, dass der untersuchte Medienobjekttyp sich in einem frühen Entwicklungsstadium befindet.

In dieser Arbeit galt die Forderung, beliebig generierte 3D-Videoobjekte miteinander vergleichen zu können. Es ist nicht ausgeschlossen, dass diese Anforderung für künftige Untersuchungen nicht mehr erforderlich ist. Wenn sich beispielsweise ein bestimmtes 3DVO-Verfahren für viele Anwendungen durchsetzt, können Qualitätsuntersuchungen auch auf der Ebene der primären oder der sekundären Repräsentation erfolgen.

Die in dieser Arbeit vorgestellte Methodik ist offen für die zweifelsfrei zu erwartende Weiterentwicklung der 3DVO-Verfahren. Die damit einhergehende Veränderung der Qualitätsaspekte kann durch angepasste und neue Qualitätsparameter berücksichtigt werden. Mit der Weiterentwicklung der 3DVO-Verfahren ist eine Verbesserung der Qualität zu erwarten. Wenn die Anzahl der Störungstypen kleiner wird und die Störungen in ihrem Ausmaß kleiner werden, lässt sich die Qualität mittels mathematischer Modelle genauer beschreiben. Auch die subjektive Bewertung wird weniger differieren, sodass die Prädiktion genauer werden kann. Damit kann ein 3DVQM für einen breiten praktischen Einsatz entwickelt werden.

Einige Qualitätsaspekte und Störungen werden in der vorliegenden Arbeit über ihre statistischen Auswirkungen erkannt. Es sind aber mathematische Modelle für konkrete Störungen (z. B. Okklusionen, Epipolarverzerrungen) erforderlich, um eine genauere Prädiktion der Qualität zu ermöglichen. Die Verbesserung solcher Modelle eröffnet ein großes Feld für künftige Forschung und Entwicklung.

Für den Einsatz des 3D-Videoobjektqualitätsmaßes in der Praxis muss die Bestimmung der Qualitätsparameter mit einem vertretbaren Aufwand möglich sein⁴. Daher sind

⁴Für die Bestimmung der in dieser Arbeit verwendeten Qualitätsparameter benötigt man beispielsweise

Untersuchungen zur Optimierung der Qualitätsparameter sowie deren rechentechnischer Umsetzung notwendig.

6.3 Schlussbetrachtung

Natürliche 3D-Videoobjekte sind ein neuer Medienobjekttyp, dessen Generierung die Entwickler noch vor viele Probleme stellt. Die Qualitätsbewertung solcher Objekte wirft viele Fragen auf, da sowohl die sich aus der Nutzung ergebenden Qualitätsanforderungen als auch die Störungen dieser Objekte sich nicht vollständig erfassen lassen. Andererseits ist die Qualitätsbewertung von 3D-Videoobjekten für die Verbesserung ihrer Generierungsverfahren unbedingt erforderlich. In der vorliegenden Arbeit wurde ein erster Ansatz zur objektiven Qualitätsbewertung solcher Objekte unabhängig von deren Generierungsverfahren vorgestellt. Aus dieser Arbeit konnte selbstverständlich kein Maß hervorgehen, mit dem es möglich ist, sehr gute Prädiktionen der subjektiven Qualitätsbewertung für alle jetzigen und künftigen 3D-Videoobjekttypen zu erzielen. Umso wichtiger war es, eine Methodik zu entwickeln und diese nachzuweisen, welche für die Anpassung an die künftige Anforderungen genutzt werden kann.

mit einem PC (Prozessortakt: 2 GHz) für ein etwa vier Sekunden dauerndes Videoobjekt etwa drei Stunden. Die Implementierung der Algorithmen erfolgte hierfür in der Programmiersprache C.

Abbildungsverzeichnis

2.1	Kette von der Akquisition bis zur Visualisierung von 2D-Bewegtbild	3
2.2	Radiale und tangentielle Verzerrungen der Linsenoptik	4
2.3	Kameramodell	5
2.4	MTF eines Fernsehkameraobjektives (nach [Wen88])	6
2.5	3D-Abtastschema	7
2.6	Bildbetrachtung: Betrachterabstand, Bildhöhe, Öffnungswinkel	8
2.7	CIE- u' , v' -Farbtafel	10
2.8	Betrachtungsabstand zur Bewertung von Fernsehbildern nach [ITU02]	12
2.9	DSCQS-Skala	15
2.10	Schritte zur Bestimmung des VQM; nach [Wol02]	22
2.11	Qualitätsmodell nach Fechter; [Fec98]	25
2.12	Bestimmung der JND-Maps	27
2.13	Methoden zur technischen Bildqualitätsmessung nach [Fec99]	31
3.1	Zwei Ansichten auf ein SNHC-Szenario mit einem 3D-Videoobjekt	33
3.2	Die Interpolation (Mitte) zwischen zwei aufgenommenen Ansichten (links und rechts)	34
3.3	Die Epipolar-Geometrie zweier Abbildungen	35
3.4	Rektifizierung	35
3.5	Trifokale Ansicht; nach [Fau01]	36
3.6	Bild einer Geraden in der 3. Ansicht; nach [Fau01]	37
3.7	Objektansicht mit zugehöriger Tiefenkarte; [Haa04]	38
3.8	Generierung der Visual Hull aus den Silhouetten	39
3.9	Morphing nach [Sei96] in drei Schritten	40
3.10	Lineares Morphing von Bildpunkten	40
3.11	Bildung der I3D-Repräsentation; nach [Ohm99a]	41
3.12	View-Synthesis der I3D-Repräsentation; nach [Ohm99a]	41
3.13	Light-Field-Rendering: Light Slab (links), Feld der Lichtstrahlen zwischen 2 Flächen (rechts)	42
3.14	Bestimmung einer blickpunktabhängigen Textur; nach [Deb96]	44
3.15	Voxel-Colouring	45
3.16	Verfahren zur Generierung von 3D-Videoobjekten; nach [Pol00]	46
3.17	Kriterien für die Generierung von 3D-Videoobjekten	47
3.18	Modell der Generierung von 3D-Videoobjekten; nach [Rit03], [Rit04]	47
4.1	Beispiel für eine 2D-Videoqualitätsbewertung	50
4.2	Vergleich von 3D-Videoobjekten	51
4.3	Referenz für 3D-Videoobjekte	51
4.4	Referenz-3DVO (links) und 3DVO mit Perspektivenfehler (rechts)	53
4.5	Silhouette eines Referenz-3DVO (links) und eine gestörte Silhouette (rechts)	54
4.6	Referenz-3DVO (links) und 3DVO mit Warping-spezifischen Fehlern (rechts)	54

4.7	Referenz-3DVO (links) und 3DVO mit texturierungsspezifischen Fehlern (rechts)	55
4.8	DFT-Koeffizienten für verschiedene Konturen	64
4.9	Referenz-3DVO (links), 3DVO mit Skalierungsfehler (Mitte) und korrigiertes 3DVO (rechts)	66
4.10	Referenz-3DVO (links), 3DVO mit Blickwinkelfehler (Mitte) und 3DVO mit Block-Matching-Korrektur (rechts)	69
4.11	Standardabweichungen des örtlichen Informationsgehaltes für Sequenzen mit fehlerhaftem Blickwinkel	70
4.12	Silhouette eines Referenz-3DVO (links) und eines gestörten 3DVO (rechts)	73
4.13	DFT-Koeffizienten der Kontur für verschiedene Fehlertypen	74
4.14	Örtl. Informationsgehalt für zwei Sequenzen mit Texturierungsfehlern	78
4.15	Verlauf des Histogrammvergleichsindex für zwei Sequenzen mit Texturierungsfehlern	80
5.1	Abfolge zur Bestimmung des 3D-Videoobjektqualitätsmaßes	87
5.2	Bildung der Qualitätsparameter (Beispiele)	88
5.3	Sichtbarkeit eines Translationsfehlers bei räumlicher Begrenzung	90
5.4	Bestimmung der Koeffizienten für das 3DVQM	99
5.5	Formular für die erweiterte subjektive Bewertung	101
5.6	Subj. Qualitätsbewertung: DMOS und Intervall der Einzelbewertungen	107
5.7	Anteile der subjektiven Bewertung für vier Testobjekte	108
5.8	Anteile α_i und $\tilde{\alpha}_i$ der subjektiven und der objektiven Bewertung	111
5.9	Streudiagramm für den DMOS und dessen Prädiktion	112
5.10	Streudiagramm für den DMOS und dessen optimierte Prädiktion	114

Tabellenverzeichnis

2.1	Festlegungen der Primärvalenzen nach EBU und FCC	10
2.2	Bewertungsskala nach [ITU02]	14
2.3	PSNR-Werte für einige Störungen	19
2.4	Wahrscheinlichkeiten der Unterscheidung für das JND-Maß [Jan02]	26
2.5	Erläuterungen zu JND-Werten [Sar01]	26
4.1	Lineare Regression von Skalierungsfehlern und deren Bewertung	65
4.2	Merkmalsmaß $w_{trans,BM-h}$ für verschiedene Fehlertypen	68
4.3	Merkmalsmaße $w_{trans,BM-0}$ und $w_{trans,kontur}$ für verschiedene Fehlertypen .	68
4.4	Merkmalsmaß $w_{skal,kontur}$ für verschiedene Fehlertypen	69
4.5	Blockbezogener Vergleich des örtlichen Informationsgehaltes	71
4.6	Blockbezogener Histogrammvergleich	71
4.7	Merkmalsmaße $w_{blockSI}$ und $w_{blockHI}$ für verschiedene Fehlertypen	73
4.8	Merkmalsmaß $w_{silhouette,DFT}$ für verschiedene Fehlertypen	75
4.9	Fehlerfälle für Silhouettenöffnungen	76
4.10	Merkmalsmaß w_{SI} für verschiedene Fehlertypen	79
4.11	Merkmalsmaß w_{HI} für verschiedene Fehlertypen	79
4.12	Beispiele für dynamische Merkmalsmaße	82
4.13	Werte der dynamische Merkmalsmaße für verschiedene Testobjekte	83
5.1	Gegenüberstellung von Qualitätsaspekten, -merkmalen und 3DVO-Qualitätsparametern	97
5.2	Initialwerte für die Gewichtungskoeffizienten	106
5.3	Qualitätsparameter für das 3DVO „IAVAS“	109
5.4	Qualitätsparameter für die 3DVO „Lara“ und „MPI-Dancer“	110
5.5	Korrelation zwischen den objektiven und subjektiven Bewertungsanteilen .	110
5.6	Initialwerte und modifizierte Werte der Gewichtungskoeffizienten	113
5.7	Korrelation zwischen den objektiven und subjektiven Bewertungsanteilen nach der Modifikation der Gewichtungskoeffizienten	113
5.8	Verbesserung der Korrelation nach der Optimierung der Gewichtungskoeffizienten	114

Verzeichnis der verwendeten Symbole

$\#$	Kardinaloperator (bestimmt die Anzahl)
$ \dots $	absoluter Betrag
$[\dots]$	ganzer Anteil (Integer-Funktion)
α	Anteil an der subjektiven Bewertung
$\tilde{\alpha}$	Anteil an der objektiven Bewertung
β	Gewichtungskoeffizient eines Qualitätsparameters
χ	Methode
λ	Wellenlänge des Lichtes
Θ, Φ	Richtung in sphärischen Koordinaten
Ψ	Qualitätsmerkmal
c/ph	cycles per degree
C	optisches Zentrum, Brennpunkt
d_{ref}	Referenzmaß (z. B. Weite der Projektionsfläche)
deg	degree (Grad)
D	Abtastraster
E	Bildebene
f	Frequenz, Ortsfrequenz
F	Fundamentalmatrix
g	Wichtungskoeffizient für die sukzessive Approximation
j	imaginäre Einheit ($\sqrt{-1}$)
\mathbf{l}_m	Epipolarlinie
L	Leuchtdichte
\mathbf{m}	Abbildung eines Punktes
M	abzubildender Punkt
N	Anzahl (der Bildpunkte)
p_i	Qualitätsparameter
P	Anzahl (der Punkte)
r	Korrelationskoeffizient nach Pearson
r_s	Rangkorrelationskoeffizient nach Spearman
R	Reife des Qualitätsmaßes
t	Zeit
T	Zeitpunkt
\mathcal{T}	Tensor
$u, v; u', v'$	Koordinaten der CIE-Farbtafeln
V_X, V_Y, V_Z	Punkt im Raum mit den Koordinaten x, y, z
$w_{\Psi, \chi}$	Maß für ein Qualitätsmerkmal Ψ
x, y, z	kartesische Koordinaten, Koordinaten des xyz-Farbraumes
Y, C_R, C_B	Koordinaten des Farbraumes gemäß ITU-R BT.601

Verzeichnis der verwendeten Abkürzungen

2D	zweidimensional
3D	dreidimensional
3DVO	dreidimensionales Videoobjekt
3DVQM	3D-Videoobjektqualitätsmaß
AB	added background (points)
ACR	absolute category rating
ANSI	American National Standards Institute
AOI	Area of Interest
CCD	charge coupled device
CIE	Commission Internationale de l'Eclairage
CRT	cathode ray tube
DCR	degration category rating
DCT	Diskrete Cosinus-Transformation
DFT	Diskrete Fourier-Transformation
DMOS	differential mean opinion score
DSCQS	double-stimulus continuous quality-scale
DSIS	double-stimulus impairment scale
DVB	Digital Video Broadcasting
EBU	European Broadcasting Union
FCC	Federal Communications Commission
FR	full reference
FTV	Free Viewpoint TV
FVV	Free Viewpoint Video
HDTV	high definition television
HI	histogram intersection
HRC	hypothetical reference circuit
HSV	hue, saturation, value (Farbraum)
HVS	human visual system
I3D	Incomplete 3D
IAVAS	Interaktive audiovisuelle Anwendungssysteme
IBR	Image-Based Rendering
ISO	International Organization for Standardization
ITS	Institute for Telecommunication Services
ITU	International Telecommunication Union
JND	just noticeable difference
LCD	liquid crystal display
LDI	layered depth images
LIDAR	light detection and ranging
Lp	Linienpaare
MAD	mean absolute difference
MAV	mean annoyance value

MF	missing foreground (points)
MOS	mean opinion score
MPEG	Moving Picture Experts Group
MSE	mean squared error
MTF	modulation transfer function
NR	no reference
OBE	object-based evaluation
PC	pair comparison
PC	Personal Computer
pel	picture element
PSNR	peak signal-to-noise ratio
PVD	preferred viewing distance
RCA	root cause analysis
RGB	Rot Grün Blau
RMS	root mean square
RR	reduced reference
SAD	sum of absolute differences
SDSCE	simultaneous double stimulus for continuous evaluation
SI	spatial (perceptual) information
SMPTE	Society of Motion Picture and Television Engineers
S/N	signal-to-noise ratio
SNHC	Synthetic/Natural Hybrid Coding
SQM	Spatial Quality Measure
SRC	source sequence
SS	single stimulus
SSCQE	single stimulus continuous quality evaluation
SSD	sum of squared differences
SSMR	single stimulus with multiple repetition
TI	temporal (perceptual) information
TFT	thin-film transistor (LCD-Technologie)
TV	Television (Fernsehen)
VGC	variation of gravity center
VQEG	Video Quality Experts Group
vQM	variation of SQM
VQM	video quality metric
VRML	Virtual Reality Modeling Language

Quellenverzeichnis

- [Ahr98] Ahrens, N. „Objektives Messen der Bildqualität“, *FKT Film- und Kino-Technik*, Hühig GmbH, Heidelberg, 52. Jahrgang, S. 331–337, Juli 1998.
- [ANSI95] American National Standards Institute ANSI Committee T1 - Telecommunications, *ANSI T1.801.01-1995: Digital Transport of Video Teleconferencing/Video Telephony Signals – Video Test Scenes for Subjective and Objective Performance Assessment*, 1995.
- [ANSI96a] American National Standards Institute ANSI Committee T1 - Telecommunications, *ANSI T1.801.02-1996: Digital Transport of Video Teleconferencing/Video Telephony Signals – Performance Terms, Definitions, and Examples*, 1996.
- [ANSI96b] American National Standards Institute ANSI Committee T1 - Telecommunications, *ANSI T1.801.03-1996: Digital Transport of One-Way Video Signals – Parameters for Objective Performance Assessment*, 1996.
- [ANSI01] American National Standards Institute ANSI Committee T1 - Telecommunications, *Video Normalization Methods Applicable to Objective Video Quality Metrics Utilizing a Full Reference Technique*, Draft Technical Report, Tucson (USA), 2001.
- [Avi98] Avidan, S.; Shashua, A.: „Novel View Synthesis by Cascading Trilinear Tensors“, *IEEE Transactions on Visualization and Computer Graphics*, Ausgabe 4, Nr. 4, S. 293–306, Oktober–Dezember 1998.
- [Bær98] Bærentzen, A. „Octree-Based Volume Sculpting“, *Proceedings of IEEE Visualization '98*, Research Triangle Park, North Carolina (USA), S. 9–12, Oktober 1998.
- [Bon96] Bonse, T. *Zur Konzeption einer visuell angepassten Beschreibung und Darstellung von Bewegtbildern*, Dissertation an der Universität Dortmund, Shaker Verlag, Aachen, 1996.
- [Bue01] Buehler, C.; Bosse, M.; McMillan, L.; Gortler, S.; Cohen, M. „Unstructured Lumigraph Rendering“, *Proceedings of ACM SIGGRAPH'01*, S. 425–432, 2001.
- [Car03] Carranza, J.; Theobalt, C.; Magnor, M.; Seidel, H.-P. „Free-Viewpoint Video of Human Actors“, *Proceedings of ACM SIGGRAPH'03*, S. 569–577, 2003.
- [Che93] Chen, S. E.; Williams, L. „View Interpolation for Image Synthesis“, *Proceedings of ACM SIGGRAPH'93*, 1993.
- [Che00] Cheung, G.; Kanade, T. „A Real Time System for Robust 3D Voxel Reconstruction of Human Motions“, *Proceedings of Computer Vision and Pattern Recognition*, S. 714–720, 2000.

- [Che02] Chen, W.-C.; Bouguet, J.-Y.; Chu, M.; Grzeszczuk, R. „Light Field Mapping: Efficient Representation and Hardware Rendering of Surface Light Field“, *Proceedings of ACM SIGGRAPH'02*, S. 447–456, 2002.
- [Cor01] Correia, P. L.; Pereira, F. „Standalone Objective Evaluation of Segmentation Quality“, *Proceedings of Workshop on Image Analysis for Multimedia Interactive Services WIAMIS*, Tampere (Finnland), Mai 2001.
- [Cor03] Correia, P. L.; Pereira, F. „Objective Evaluation of Video Segmentation Quality“, *IEEE Transactions on Image Processing*, Ausgabe 12, S. 186–200, Februar 2003.
- [Deb96] Debevec, P. E.; Taylor, C. J.; Malik, J. „Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach“, *Computer Graphics*, Ausgabe 30, Annual Conference Series, S. 11–20, 1996.
- [Dru02] Drumm, H.; Kühhirt, U.; Rittermann, M.; Reiter, U. „Anwendungssysteme für MPEG-4“, *Jahrestagung der Fernseh- und Kinotechnischen Gesellschaft*, Zürich (Schweiz), 2002.
- [Due03] Duelli, M. *Evaluation und Implementierung von Verfahren zur View-Synthese im IAVAS-Kontext*, Diplomarbeit an der Technischen Universität Ilmenau, Ilmenau, 2003.
- [EBU75] European Broadcasting Union EBU, *Tech. 3213-E: Standard for Chromaticity Tolerances for Studio Monitors*, 1975.
- [Eis99] Eisert, P.; Steinbach, E.; Girod, B. „Multi-Hypothesis Volumetric Reconstruction of 3-D Objects from Multiple Camera Views“, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing ICASSP'99*, Phoenix (USA), S. 3509–3512, März 1999.
- [Erd01] Erdem, Ç.; Tekalp, A.; Sankur, B. „Metric for Performance Evaluation of Video Segmentation and Tracking Without Ground-Truth“, *Proceedings of International Conference on Image Processing ICIP'2001*, Thessaloniki (Griechenland), 2001.
- [Fau93] Faugeras, O. *Three-Dimensional Computer Vision, A Geometric Viewpoint*, The MIT Press, Cambridge (USA), 1993.
- [Fau01] Faugeras, O.; Luong, Q.-T. *The Geometry of Multiple Images*, Massachusetts Institute of Technology, The MIT Press, Cambridge MA (USA), 2001.
- [Fec98] Fechter, F. „Objektive Beurteilung der Qualität komprimierter Bildfolgen“, *FKT Film- und Kino-Technik*, Hüthig GmbH, Heidelberg, 52. Jahrgang, S. 417–421, Juli 1998.
- [Fec99] Fechter, F. „Technische Bildqualitätsmessung – Ein Vergleich alternativer Methoden“, *ITG-Fachbericht „Elektronische Medien: Technologien, Systeme, Anwendungen“*, *Vorträge auf dem 8. Dortmunder Fernsehseminar*, Dortmund, S. 53–62, 1999.
- [Fec00] Fechter, F. *Zur Beurteilung der Qualität komprimierter Bildfolgen*, Habilitationsschrift an der Technischen Universität Carolo-Wilhemina zu Braunschweig, VDI Verlag GmbH, Düsseldorf, 2000.

- [Fuj96] Fujii, T.; Kimoto, T.; Tanimoto, M. „Ray Space Coding for 3D Visual Communication“, *Proceedings of Picture Coding Symposium PCS'96*, S. 447–451, März 1996.
- [Gel04] Gelasca, E. D.; Ebrahimi, E.; Farias, M. C. Q.; Mitra, S. K. „Impact of Topology Changes in Video Segmentation Evaluation“, *Proceedings of the 5th Workshop on Image Analysis for Multimedia Interactive Services WIAMIS*, Lissabon (Portugal), April 2004.
- [Gre03] Grewatsch, S.; Müller, E. „Kompression von Tiefenkarten für das dreidimensionale Fernsehen“, *4. IuK-Tage Mecklenburg-Vorpommern*, Rostock, 18.–20. Juli, 2003.
- [Gu99] Gu, J. „3D Wavelet-Based Video Codec with Human Perceptual Model“, Master's Thesis an der University of Maryland (USA), 1999.
- [Haa04] von der Haar, J. *Erweiterte Blickpunktwahl für die View-Synthese von 3D-Videoobjekten*, Diplomarbeit an der Technischen Universität Ilmenau, Ilmenau, 2004.
- [Haf98] Hafner, W. *Segmentierung von Video-Bildfolgen durch Adaptive Farbklassifikation*, Dissertation an der Technischen Universität München, München, 1998.
- [Hau94] Hauske, G. *Systemtheorie der visuellen Wahrnehmung*, B. G. Teubner, Stuttgart, 1994.
- [Her03] Herzog, N. *Untersuchung und Evaluierung von Verfahren zur 3D-Videoerzeugung*, Diplomarbeit an der Technischen Universität Ilmenau, Ilmenau, 2003.
- [Hol94] Hollmann, T. *Zur Gewinnung von Bewegungsinformationen in einer Kamera mit hoher Bildaufnahmefrequenz*, Dissertation an der Universität Dortmund, Shaker Verlag, Aachen, 1994.
- [Isi02] Isidoro, J.; Sclaroff, S. „Stochastic Mesh-Based Multiview Reconstruction“, *Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission*, S. 568–577, 2002.
- [ISO98] International Organization for Standardization ISO/IEC JTC 1/SC 29/WG 11, *ISO 14496 Information Technology – Generic Coding of Audio-Visual Objects (MPEG-4)*, 1998pp.
- [ISO03a] International Organization for Standardization ISO/IEC JTC 1/SC 29/WG 11; Smolić, A.; Kimata, H. *N5877: Applications and Requirements for 3DAV*, Trondheim (Norwegen), Juli 2003.
- [ISO03b] International Organization for Standardization ISO/IEC JTC 1/SC 29/WG 11; Smolić, A.; Kimata, H. *N5959: Description of Exploration Experiments in 3DAV*, Brisbane (Australien), Oktober 2003.
- [ITU94a] International Telecommunication Union ITU, *Recommendation ITU-R BT.601-4: Encoding Parameters of Digital Television for Studios*, 1994.
- [ITU94b] International Telecommunication Union ITU, *Recommendation ITU-R BT.802-1: Test Pictures and Sequences for Subjective Assessments of Digital Codecs Conveying Signals Produced According to Recommendation ITU-R BT.601*, 1994.

- [ITU99] International Telecommunication Union ITU, *Recommendation ITU-T P.910: Subjective Video Quality Assessment Methods for Multimedia Applications*, 1999.
- [ITU00] International Telecommunication Union ITU, *Recommendation ITU-T P.920: Interactive Test Methods for Audiovisual Communications*, 2000.
- [ITU01] International Telecommunication Union ITU, *Recommendation ITU-T J.144: Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference*, 2001.
- [ITU02] International Telecommunication Union ITU, *Recommendation ITU-R BT.500-11: Methodology for the Subjective Assessment of the Quality of Television Pictures*, 2002.
- [Jan02] Janssen, J.; Coppens, T.; De Vleeschauwer, D. „Quality Assessment of Video Streaming in the Broadband Era“, *Proceedings of Advanced Concepts for Intelligent Vision Systems ACIVS 2002*, Ghent (Belgien), S. 38–45, 2002.
- [Kay96] Kayser, A. *Modellierung der Bildqualität datenkomprimierter stereoskopischer Fernsehbilder*, Dissertation an der Technischen Universität München, VDI Verlag GmbH, Düsseldorf, 1995.
- [Küh03] Kühhirt, U.; Rittermann, M. „Implementierung des Objekt- und 3D-Szenenkonzeptes von MPEG-4“, *48. Internationales Wissenschaftliches Kolloquium der TU Ilmenau*, Ilmenau, September 2003.
- [Lam04] Lamboray, E.; Waschbüsch, M.; Würmlin, S.; Gross, M.; Pfister, H. „Unconstrained Free-Viewpoint Video Coding“, angekündigt für: *IEEE International Conference on Image Processing ICIP 2004*, Singapur (Singapur), Oktober 2004.
- [Lau94] Laurentini, A. „The Visual Hull Concept for Silhouette Based Image Understanding“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Ausgabe 16, S. 150–162, Februar 1994.
- [Lev96] Levoy, M.; Hanrahan, P. „Light-Field Rendering“, *Proceedings of ACM SIGGRAPH'96*, 1996.
- [Lue01] Luebke, D.; Hallen B. „Perceptually Driven Simplification for Interactive Rendering“, *Rendering Techniques*, Springer-Verlag, London (Großbritannien), Juni 2001.
- [Mar02] Marziliano, P.; Dufaux, F.; Winkler, S.; Ebrahimi, T. „A No-Reference Perceptual Blur Metric“, *Proceedings of International Conference on Image Processing ICIP 2002*, Ausgabe 3, S. 57–60, Rochester NY (USA), September 2002.
- [Mas04] Masry, M.; Hemami, S. S. „A Metric for the Continuous Evaluation of Compressed Video with Severe Distortions“, *Signal Processing: Image Communication, Special issue on Video Quality*, Ausgabe 19, Elsevier B. V., S. 133–146, Februar 2004.
- [Mat00] Matusik, W.; Buehler, C.; Raskar, R.; Gortler, S.; McMillan, L. „Image Based Visual Hulls“, *Proceedings of ACM SIGGRAPH'00*, 2000.

- [Mat04] Matsuyama, T.; Wu, X.; Takai, T.; Wada, T. „Real-Time Dynamic 3-D Object Shape Reconstruction and High-Fidelity Texture Mapping for 3-D Video“, *IEEE Transactions on Circuits and Systems for Video Technology*, Ausgabe 14, S. 357–369, März 2004.
- [Mäu95] Mäusl, R. *Fernsehtechnik*, Hüthig Buch Verlag, Heidelberg, 1995.
- [McMil95] McMillan, L.; Bishop, G. „Plenoptic Modeling: An Image-Based Rendering System“, *Proceedings of ACM SIGGRAPH'95*, 1995.
- [Mik01] Mikic, I.; Triverdi, M.; Hunter, E.; Cosman, P. „Articulated Body Posture Estimation from Multicamera Voxel Data“, *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai Marriott HI (USA), Dezember 2001.
- [Moe97] Moezzi, S.; Tai, L.; Gerard, P. „Virtual View Generation for 3D Video“, *IEEE Multimedia*, Ausgabe 4, S. 18–26, Januar 1997.
- [Müh02] Mühlmann, K. *Design und Implementierung eines Systems zur schnellen Rekonstruktion dreidimensionaler Modelle aus Stereobildern*, Dissertation an der Universität Mannheim, Mannheim, 2002.
- [Mül99] Müller, K.; Ohm, J.-R. *Wavelet-Based Contour Descriptor, Moment-Based Contour Descriptor*, MPEG-7 Proposal No. P567/P568, Februar 1999.
- [Mül04] Müller, K.; Smolić, A.; Kaspar, B.; Merkle, P.; Rein, T.; Eisert, P.; Wiegand, T. „Octree Voxel Modeling with Multi-View Texturing“, *Proceedings of the 5th Workshop on Image Analysis for Multimedia Interactive Services WIAMIS*, Lissabon (Portugal), April 2004.
- [Mys01] Myszkowski, K.; Tawara, T.; Akamine, H.; Seidel, H.-P. „Perception-Guided Global Illumination Solution for Animation Rendering“, *Proceedings of ACM SIGGRAPH'01*, S. 221–230, 2001.
- [Nae00] Naemura, T.; Harashima, H. „Ray-based approach to Integrated 3D Visual Communication“, *SPIE Three-Dimensional Video and Display: Devices and Systems (Critical Reviews of Optical Science and Technology)*, Ausgabe CR76, S. 282–305, November 2000.
- [Ohm95] Ohm, J.-R. *Digitale Bildcodierung*, Springer-Verlag, Berlin, 1995.
- [Ohm97a] Ohm, J.-R.; Müller, K. „Incomplete 3D Representation of Video Objects for Multiview Applications“, *Proceedings of Picture Coding Symposium PCS'97*, Berlin, S. 427–432, September 1997.
- [Ohm97b] Ohm, J.-R.; Höhne, H. D. „Disparitätsgesteuerte Key-Mischung stereoskopischer Videosignale“, *ITG-Fachbericht „Elektronische Medien: Technologien, Systeme, Anwendungen“*, Vorträge auf dem 7. Dortmunder Fernsehseminar, Dortmund, S. 179–184, 1997.

- [Ohm99a] Ohm, J.-R.; Müller, K. „Incomplete 3D – Multiview Representation of Video Objects“, *IEEE Transactions on Circuits and Systems for Video Technology; Special Issue on SNHC*, S. 389–400, März 1999.
- [Ohm99b] Ohm, J.-R. *Bildverarbeitung 2*, Skript zur Vorlesung, Technische Universität Berlin, Berlin, 1999.
- [deOli00] de Oliveira Neto, M. M. *Relief Texture Mapping*, Dissertation an der University of North Carolina at Chapel Hill, Chapel Hill (USA), 2000.
- [Pol98] Pollard, S.; Hayes, S. „View Synthesis by Edge Transfer with Application to the Generation of Immersive Video Objects“, *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, ACM Press/ACM SIGGRAPH, New York (USA), S. 335–342, 1998.
- [Pol00] Pollefeys, M.; Kang, S. B.; Slabaugh, G. C.; Cornelis, K. „Panel Session on Visual Scene Representation“, *3D Structure from Images - SMILE 2000: Second European Workshop on 3D Structure from Multiple Images of Large-Scale Environments, Dublin, Ireland, July 2000*, Springer Verlag, Heidelberg, S. 161–169, 2000.
- [Rav98] Ravel, M.; Lubin, J.; Schertz, A. „Prüfung eines Systems für die objektive Messung der Bildqualität durch den Vergleich von objektiven und subjektiven Testergebnissen“, *FKT Film- und Kino-Technik*, Hüthig GmbH, Heidelberg, 52. Jahrgang, S. 567–575, Oktober 1998.
- [Red00] Redert, A. *Multi-Viewpoint Systems*, Dissertation an der TU Delft (Niederlande), P. A. Redert, 2000.
- [Rit03] Rittermann, M. „3D-Videoobjekte für immersives TV“, *ITG-Fachbericht „Elektronische Medien: Technologien, Systeme, Anwendungen“, Vorträge auf dem 10. Dortmunder Fernsehseminar*, Dortmund, September/Oktober 2003.
- [Rit04] Rittermann, M. „A Proposal for the Quality Assessment of 3D Video Objects“, *Proceedings of the 5th Workshop on Image Analysis for Multimedia Interactive Services WIAMIS*, Lissabon (Portugal), April 2004.
- [Sar01] Sarnoff Corporation, *JND: A Human Vision System Model for Objective Picture Quality Measurements*, White Paper, Princeton NJ (USA), 2001.
- [Sar02] Sarnoff Corporation, *Measuring Image Quality, Sarnoff's JNDmetrix*, Technology Overview, Princeton NJ (USA), 2002.
- [Sch94] Schneuwly, B. *Ein Qualitätsmassstab für die Bildwiedergabe*, Dissertation an der Eidgenössischen Technischen Hochschule Zürich (Schweiz), Hartung-Gorre Verlag, Konstanz, 1994.
- [Sei96] Seitz, S. M.; Dyer, C. R. „View Morphing“, *Proceedings of ACM SIGGRAPH'96*, 1996.
- [Sei97] Seitz, S. M. *Image-Based Transformation of Viewpoint and Scene Appearance*, Dissertation an der University of Wisconsin-Madison (USA), 1997.

- [Sei99] Seitz, S. M.; Dyer, C. R. „Photorealistic Scene Reconstruction by Voxel Coloring“, *International Journal of Computer Vision*, Ausgabe 35, Nr. 2, November 1999.
- [Sha98] Shade, J.; Gortler, S.; He, L.-W.; Szeliski, R. „Layered Depth Images“, *Proceedings of ACM SIGGRAPH'98*, 1998.
- [Sik03] Sikora, T. „MPEG-4 objektbasierte Videocodierung (MPEG-4 Object-Based Video Coding)“, *it Information Technology*, Ausgabe 45, Oldenbourg Wissenschaftsverlag, München, S. 273–279, Mai 2003.
- [Smo03] Smolić, A.; Fehn, C.; Müller, K. „MPEG 3DAV – Video-Based Rendering for Interactive TV Applications“, *ITG-Fachbericht „Elektronische Medien: Technologien, Systeme, Anwendungen“*, Vorträge auf dem 10. Dortmunder Fernsehseminar, Dortmund, September/Okttober 2003.
- [Smo04] Smolić, A.; Müller, K.; Merkle, P.; Rein, T.; Kautzner, M.; Eisert, P.; Wiegand, T. „Free Viewpoint Video Extraction, Representation, Coding, and Rendering“, angekündigt für: *IEEE International Conference on Image Processing ICIP 2004*, Singapur (Singapur), Oktober 2004.
- [SMPTE03] Society of Motion Picture and Television Engineers SMPTE, *Recommended Practice RP 222-2003: Standard Definition Evaluation Materials for Digital Television*, 2003.
- [Tei90] Teichner, D. *Fernsehen mit erhöhter Bildqualität: bildinhaltsabhängige Filterung von PAL-Signalen*, Drei-R-Verlag, Berlin, 1990.
- [Tra98] Trauberg, M. „Ein neues Verfahren für die Bildqualitätsüberwachung in MPEG-basierten Übertragungssystemen“, *FKT Film- und Kino-Technik*, Hüthig GmbH, Heidelberg, 52. Jahrgang, S. 411–416, August 1998.
- [Tur04] Turaga, D. S.; Chen, Y.; Caviedes, J. „No Reference PSNR Estimation for Compressed Pictures“, *Signal Processing: Image Communication, Special issue on Video Quality*, Ausgabe 19, Elsevier B. V., S. 173–184, Februar 2004.
- [Vil99] Villegas, P.; Marichal, X.; Salcedo, A. „Objective Evaluation of Segmentation Masks in Video Sequences“, *Proceedings of Workshop on Image Analysis for Multimedia Interactive Services WIAMIS*, Berlin, S. 85–88, Mai 1999.
- [VQEG03] Video Quality Experts Group VQEG, *Validation of Objective Models of Video Quality Assessment, Phase II*, Draft Final Report from the Video Quality Experts Group, 2003.
- [Wan02] Wang, Y., Ostermann, J., Zhang, Y.-Q., *Video Processing and Communications*, Prentice-Hall Inc., Upper Saddle River NJ (USA), 2002.
- [Wen88] Wendland, B. *Fernsehtechnik, Band 1: Grundlagen*, Dr. Alfred Hüthig Verlag, Heidelberg, 1988.
- [Wer95] Werner, T.; Hersch, R. D.; Hlaváč, V. „Rendering Real-World Objects Using View Interpolation“, *International Conference on Computer Vision*, Cambridge MA (USA), S. 957–962, Juni 1995.

- [Win00] Winkler, S. *Vision Models and Quality Metrics for Image Processing Applications*, Dissertation an der École Polytechnique Fédérale de Lausanne, Lausanne (Schweiz), 1994.
- [Wol98] Wolf, S.; Pinson, M. H. „In-Service Performance Metrics for MPEG-2 Video Systems“, *Made to Measure 98 – Measurement Techniques of the Digital Age Technical Seminar*, International Academy of Broadcasting (IAB), ITU, Technische Universität Braunschweig; Montreux (Schweiz), November 1998.
- [Wol02] Wolf, S.; Pinson, M. H. *Video Quality Measurement Techniques*, NTIA Report, Boulder (USA), 2002.
- [Wür02] Würmlin, S.; Kern, N. J.; Staadt, O. G.; Gross, M. H. *3D Video Particles*, Technical Report No. 376, Institute of Scientific Computing, Eidgenössische Technische Hochschule Zürich, Zürich (Schweiz), 2002.
- [Wür04] Würmlin, S.; Lamboray, E.; Gross, M. „3D Video Fragments: Dynamic Point Samples for Real-time Free-Viewpoint Video“, *Computers and Graphics; Special Issue on Coding, Compression and Streaming Techniques for 3D and Multimedia Data*, Ausgabe 28, Nr. 1, Elsevier Ltd., S. 3–14, 2004.
- [Zie04] Ziegler, G.; Lensch H. P. A.; Ahmed, N.; Magnor, M.; Seidel, H.-P. „Multi-Video Compression in Texture Space“, angekündigt für: *IEEE International Conference on Image Processing ICIP 2004*, Singapur (Singapur), Oktober 2004.